CONFIDENTIAL

CENTRAL INTELLIGENCE AGENCY

INFORMATION REPORT

25X1

COUNTRY International

SUBJECT Third International Biometric Conference, 1-5 Sep 53, at Bellagio, Italy.

DATE DISTR. 15 MAR 54

NO. OF PAGES 2

NO. OF ENCLS.

SUPP. TO REPORT NO.

25X1

THIS IS UNEVALUATED INFORMATION

25X1

25X1 1. The Biometric Conference was relatively small and informal

25X1

25X1 2. The list of members (a booklet containing this list and abstracts of papers presented is available at the end of this report) contains 95 names on the main list and 16 "additions".

25X1

25X1

25X1

25X1

25X1

CONFIDENTIAL

DISTRIBUTION → STATE X ARMY S60 X NAVY S/I X AIR S60 X FBI CIA/M X

25X1

Next 1 Page(s) In Document Exempt

III INTERNATIONAL BIOMETRIC CONFERENCE

Bellagio, 1-5 Sept 1953

LIST OF MEMBERS

ABSTRACTS OF PAPERS

LIST OF MEMBERS

Mr. **F.J. ANSCOMBE**	Statistical Laboratory St. Andrew's Hill **CAMBRIDGE** (England)
Dr. **G. BARBENSI**	Viale S. Lavagnini, 28 **FIRENZE** (Italia)
Prof. **C. BARIGOZZI**	Istituto di Genetica Via Celoria, 10 **MILANO** (Italia)
Dr. **N.A. BARRICELLI**	15 Edwards Place **PRINCETON N.J.** (U.S.A.)
Prof. **M.S. BARTLETT**	Dept. of Mathematics The University **MANCHESTER 13** (England)
Mr. **M.W. BENTZON**	Berberisvaenget 33 **LINGBY** (Danmark)
Dr. **C. BIAGINI**	Via B. Latini, 122 **FIRENZE** (Italia)
Mr. **R.E. BLACKITH**	Imperial College of Science and Technology Field Station, Ashurst Lodge **SUNNINGHILL, Berkshire** (England)
Dr. **C.I. BLISS**	P.O. Box 1106 - The Biometric Society **NEW HAVEN 4, Conn.** (U.S.A.)
Dr. **N. BLOMQVIST**	Inst. of Mathematical Statistics Norrtullsgatan 16 **STOCKHOLM Va.** (Sweden)
Prof. **M. BOLDRINI**	Piazza Repubblica, 23 **MILANO** (Italia)

Dr. J.O. IRWIN	London School of Hygiene and Tropical Medicine Keppel St. **LONDON W.C.1** (England)
Mr. A. KAELIN	Zederstrasse 4 **ZUERICH 7/32** (Suisse)
Dr G. KARREMAN	Committee on Mathematical Biology - (University of Chicago) 5741 Drexel Ave **CHICAGO 37, Ill.** (U S A.)
Dr. M. KEULS	Institut voor de Veredeling van Tuinbouwgewassen Heerenstraat 25 **WAGENINGEN** (The Netherlands)
Mr. E. van der LAAN	Ministry of Agriculture Room 315 **THE HAGUE** (The Netherlands)
M.me M.J. LAURENT-DUHAMEL	8 Square Vauban **VIROFLAY - Seine et Oise** (France)
Mr. F.B. LEECH	Statistics Department Rothamsted Experimental Station **HARPENDEN, Herts** (England)
Mr. J.M. LEGAY	Station de Recherches Séricicoles **ALES (GARD)** - (France) -
Dr. A. LENGER	37 Rue Colonel Chaltin, Uccle **BRUXELLES** (Belgique)
Prof. I.M. LERNER	Istituto di Genetica Piazza Botta **PAVIA** (Italia)
Prof. A. LINDER	24, Avenue de Champel **GENEVE** (Suisse)

Dr. F. **LUKACS**	3727 Van Ness Street. N W **WASHINGTON 16**, **D.C.** (U S A)
Prof. **J L. LUSH**	Animal Husbandry Dept. I S C **AMES** Iowa (U S A)
Dr. **G. MAGNI**	Istituto di Genetica Piazza Botta **PAVIA** (Italia)
Dr. **L. MARTIN**	104 Avenue Albert **BRUXELLES** (Belgique)
Prof. **K. MATHER**	Genetics Department The University Edgbaston **BIRMINGHAM 15** (England)
Prof. **G. MONTALENTI**	Istituto di Genetica Via Mezzocannone 8 **NAPOLI** (Italia)
Prof. **E. MORICE**	Directeur Ecole d'Application Institut de Statistique 1 Rue de la Banque **PARIS** (France)
Dr. **C.A.G. NASS**	Nederlands Instituut voor Praeventieve Geneeskunde Wassenaarse Weg, 56 **LEIDEN** (The Netherlands)
Mr. **J A. NELDER**	National Vegetable Research Station **WELLESBOURNE**, Warwick (England)
Prof. **O. NISSEN**	The Agricultural College of Norway **VOLLEBEKK** (Norway)
Prof. **P. OTTESTAD**	The Agricultural College of Norway **VOLLEBEKK** (Norway)

Dr. **A.R.G. OWEN**	Dept. of Genetics 44 Storeys Way **CAMBRIDGE** (England)
Dr. **A.F. PARKER-RHODES**	20, Desley-Taylor Rd, **CAMBRIDGE** (England)
Mr. **S.C. PEARCE**	East Malling Research Station **EAST MALLING, MAIDSTONE, Kent** (England)
Dr. **P. PÉLEGRIN**	Institut des Fruits et Agrumes Coloniaux 6 rue du Général Clergerie **PARIS XVI** (France)
Mr. **S. PETO**	Ministry of Supply, Microb. Res. Dept. **PORTON nr. SALISBURY, Wilts.** (England)
Dr. **A. PREVITERA**	Lab. Provinciale Igiene e Profilassi Via Matteotti **MASSA-APUANIA** (Italia)
Prof. **R. PRIGGE**	Paul-Ehrlich, Strasse 44 **FRANKFURT a/M** (Germany)
Prof. **C.R. RAO**	Statistical Laboratory Presidency College **CALCUTTA** (India)
Prof. **G. RASCH**	Skovmindeves 14 **HOLTE** (Danmark)
Dr. **R. READ**	Distiller's Co. Ltd. Great Burgh **EPSOM, Surrey** (England)

Mr. D. Van der REYDEN	Tobacco Research Board, P O. Box 1909 **SALISBURY (Southern Rhodesia)** (Africa)
Prof. S. RIOS GARCIA	C. DR. Bobillo, 17 **MADRID** (España)
Mr. A.H.L. ROTTI	Avenue Elisabeth 5 **MERELBEKE** (Belgique)
Mr. M.R. SAMPFORD	177, Godston Road, Wolvercote **OXFORD** (England)
Dr. D. SCHWARTZ	Seita, 2 Avenue d'Orsay **PARIS VII** (France)
Dr. R. SCOSSIROLI	Istituto di Genetica Piazza Botta **PAVIA** (Italia)
Dr. F. SELLA	Istituto Sieroterapico Milanese Via Darwin, 20 **MILANO** (Italia)
Dr. L.G. SILVESTRI	Istituto Superiore di Sanità Viale Regina Margherita 299 **ROMA 735** (Italia)
Dr. C.A.B. SMITH	The Galton Laboratory University College, Gower St. **LONDON W.C.1** (England)
Dr. P.V. SUKHATME	c/o FAO Viale delle terme di Caracalla **ROMA** (Italia)
Dr. L. TABAH	Institut National d'Etudes Démographiques 23 Avenue F. Roosevelt **PARIS VIII** (France)

Dr. J.M. TANNER	Sherrington School of Physiology St. Thomas's Hospital **LONDON SE I** (England)
Prof. V. TONOLLI	Istituto Italiano Idrobiologia **PALLANZA** (Italia)
Dr. A. TOVAR DE LEMOS	r. Viriato. 5 **LISBONNE** (Portugal)
Prof. L.A. VESSEREAU	41 Rue Boulard **PARIS 14^{O}** (France)
Mr. D.R. WESTGARTH	The Rubber Research Institute P.O. Box 150 **KUALA LAMPUR** (Malaya)
Prof. J. WISHART	Statistical Laboratory St. Andrew's Hill **CAMBRIDGE** (England)
Prof. H. WOLD	Statistics Institute Skolgatan 1 **UPPSALA** (Sweden)
Dr. F. YATES	Rothamsted Experimental Station **HARPENDEN, Herts** (England)

***

ABSTRACTS OF PAPERS

F.J. Anscombe (Statistical Laboratory - University of Cambridge)

FIXED-SAMPLE-SIZE ANALYSIS OF SEQUENTIAL OBSERVATIONS

Most methods in common use for the statistical analysis of data are based on the assumption that the sample size was chosen in advance before the observations were taken. In practice, the sample size is often not fixed in advance but depends on fortuitous events not connected with the observations or it may sometimes depend on the observations themselves. Very often there is little or no error in using fixed-sample-size methods of analysis in such cases, i e. in supposing that the eventual sample size was chosen in advance Some typical examples are given.

It is possible, however to devise sequential sampling rules which completely invalidate certain fixed sample-size methods of analysis. For example there is a sequential sampling rule by which one may prove (at any preassigned level of significance) that a coin is biased, and also a rule by which one may prove that the coin is not biased: both rules apply equally well whether or not the coin is in fact biased

The fixed sample-size methods of analysis which may possibly be incorrect when applied to observations obtained with a sequential sampling rule all have the common feature that they involve some statement about the probability distribution of the observations (or of some function of the observations) but do not introduce any prior probability distribution: for example, a significance test based on a critical region, or a confidence interval, or an unbiased point-estimate.

On the other hand, any method of analysis which uses the observations only in their likelihood function (such as a minimum risk decision procedure or a statement of posterior probabilities based on a prior probability distribution) is independent of the sampling rule followed, and it is always legitimate to treat the observations as if the sample size had been chosen in advance.

****

G. Barbensi (Firenze)

L'INSEGNAMENTO DELLA BIOMETRIA

Definita la biometria in base ai concetti oggi generalmente accettati considerato il contributo che al suo sviluppo viene portato dai biologi dai matematici e dagli statistici si considerano le varie fasi che quello sviluppo ha avuto in Italia per concludere con un quadro della situazione attuale soprattutto per quanto riguarda l insegnamento di quella materia nei nostri Istituti universitari

Considerata l'inadeguatezza di questo insegnamento, se ne analizzano le cause le quali, mentre in parte sono da ricercarsi nella deficienza di provvedimenti di legge per l istituzione di insegnamenti adeguati alle necessità in parte dipendono dalle difficoltà che gli studenti incontrano nel seguirli difficoltà che sono essenzialmente da attribuirsi alla mancanza dei necessari fondamenti matematici

Viene posto in evidenza l interesse dimostrato dai biometri italiani sia nella prima Riunione Italiana della Biometric Society a Milano nel 1951 sia nella Riunione a Firenze di quest'anno nella quale si procedette alla costituzione della Regione Italiana della Biometric Society per l'insegnamento della biometria interesse che si concretò nella formulazione di due inviti al Ministro della Pubblica Istruzione perchè attuasse i richiesti insegnamenti.

Viene esposto un programma di studi per studenti delle Facoltà biologiche (Scienze Naturali Scienze biologiche Scienze agrarie Medicina Medicina veterinaria Farmacia) tenendo conto del diverso grado di cultura matematica propedeutica e si considera l opportunità che i corsi siano fondamentali (obbligatorio) o complementari (facoltativi) a seconda dei casi.

Viene quindi esaminata l opportunità di rendere obbligatorio il Corso di biometria per alcune categorie di laureati perfezionandi od assistenti. Si esamina il problema della creazione di docenti in numero adeguato per concludere che. messi sulla via della realizzazione anche la nostra Istruzione Superiore potrà annoverare la biometria come materia di insegnamento nelle Facoltà biologiche, con quella estensione e sviluppo inerenti alla importanza da essa acquistata nello studio dei problemi biologici.

****

M W. **Bentzon** (Statens Seruminstitut Copenhagen)

ON THE STATISTICAL EVALUATION OF DOSE RESPONSE CURVES IN CASE THE DOSE INTERVALS ARE LARGE

Two different situations are considered· 1 · Dose response curves with quantal response and 2 · Dose response curves with quantitative response. The mean value and the variance of various estimates of the median effective dose are considered as functions of the true median effective dose, the slope of the response curve and the logarithmic dose interval. (The latter are taken to be equal over the whole dose range). When the dose intervals are small the ordinary estimates of the median effective dose esually is unbiased and the variance depends upon the slope × interval product only. This rule however breaks down as the intervals are increased the estimates and the variances becoming dependent upon the location of the true median effective dose within a dose interval. This effect is investigated in some cases of practical interest.

****

C.I. Bliss (Yale University)

A COURSE IN BIOMETRY FOR GRADUATE STUDENTS IN BIOLOGY

Experience gained in teaching biometry at Yale University to graduate students majoring in different branches of biology is reviewed. Originally two courses were offered in alternate years or terms, differing partly in statistical content and approach but primarily in illustrative examples, one for pharmacologists and bacteriologist and the other for botanists and zoologists. Later these were combined into a single class with the addition of a large proportion of forestry majors. The class has averaged about nine students, consisted of two or three one-hour lectures a week for one or two semesters and required from 12 to 15 laboratory hours per week in the solution of numerical examples. The course is introductory, involves a minimum of mathematics, and considers the logic of experimental design and statistical analysis with the objective of enabling potential biological investigators to design and evaluate their own research more effectively and intelligently. A detailed outline has been developed through the years and serves as the main text material How well the course has accomplished its aims and how it might be improved are considered in the light of a poll of all students who have taken the course.

****

N. Blomqvist (Institute of Mathematical Statistics Stockholm)

RANK ANALYSIS OF INCOMPLETE BLOCK DESIGNS

Suppose that t treatments shall be compared in an experiment where the number of experimental units k that are available in a block is less than t. This situation naturally leads into an incomplete block design. Led b be the number of blocks used.

In order to test the observed differences between the treatment effects we rank the observations in each block from 1 to k according to their order of magnitude and compute the rank sum S_j for each treatment Since each treatment occurs equally often (r times where $r.t = b.k$) the expected value of S_j under the null hypothesis (no treatment effects) is $r.\frac{k+1}{2}$. A natural test statistic is then

$$\sum_{j=1}^{t} (S_j - r.\frac{k+1}{2})^2$$

the distribution of which can be computed. It can be proved that the limiting distribution of

$$\frac{1}{b} \cdot \frac{12(t-1)}{k^3-k} \sum_{j=1}^{t} (S_j - r.\frac{k+1}{2})^2$$

is a χ^2 - distribution with t-1 d.f. as $b \to \infty$.

****

W.G. Cochran (Johns Hopkins University - Baltimore, Md.)

THE COMBINATION OF ESTIMATES FROM DIFFERENT EXPERIMENTS

In a series of experiments, each designed to estimate the same quantity μ, the i^{th} experiment provides an estimate x_i of μ and an estimate s_i^2 of the variance of x_i, based on n_i degrees of freedom. The experiments may be, for instance, determinations of a physical constant by different scientists, or bioassays conducted in different laboratories, or agricultural field experiments carried out in different parts of a region.

This paper discusses the problem of making a combined estimate of μ. The best combined estimate depends on the nature of the data. The first step in the analysis is to determine whether the x_i agree with each other within the limits of their experimental errors s_i^2. In practice, owing to differences in experimental techniques, presence of biases, or a real variation in μ from one experiment to another, the x_i frequently do not agree. Various implications of this situation are discussed; the recommended estimates of μ are either the unweighted mean $\bar{x}$, or a semiweighted mean

$$\bar{x}_{sw} = \Sigma \frac{x_i}{s_\mu^2 + s_i^2} \Big/ \Sigma \frac{1}{s_\mu^2 + s_i^2} ,$$

where s_μ^2 is an estimate of the variation in μ from experiment to experiment.

If there is no real variation in the x_i from one experiment to another, the recommended estimates are the unweighted mean, or the weighted mean

$$\bar{x}_w = \Sigma \frac{x_i}{s_i^2} \Big/ \Sigma \frac{1}{s_i^2} ,$$

or a partially-weighted mean. Specific recommendations are given about the use of each type of mean with illustrations for actual data.

****

D J. Finney (University of Oxford)

FUNCTIONAL RELATIONSHIPS IN EXPERIMENTATION
(Their Role in the Design and Analysis of Experiments)

Agricultural and biological experiments may be divided into three categories in respect of the extent to which functional relationships are important:

1) Experiments in which such relationships are irrelevant or even non-existent.
2) Experiments for which an underlying structure of functional relationships between the effects of different treatments is apparent but only certain characteristics are directly of interest and other details of the relationship matter little.
3) Experiments whose primary object is the study of a functional relationship.

This classification is an operational convenience rather than a clear-cut separation.

Examples of experiments falling into these categories will be presented. Category 1) clearly contains little of interest for the present purpose and 3) occurs relatively infrequently. Evidence will be presented that for category 2) in a well-designed experiment characteristics of the functional relationship other than those under study often do not materially affect the validity of estimation: for example, the position and magnitude of a maximum on a curve may be estimated fairly satisfactorily without knowledge of the precise form of the curve, provided that observations have been made on points well-distributed about the maximum. Nevertheless, careful utilization of all existing information on a functional relationship at the time of planning a new experiment may help greatly in the obtaining of high precision for new estimates.

****

Sir Ronald Fisher (University of Cambridge)

THE VARIABILITY IN THE LENGTH OF GERM PLASM STILL HETEROGENEOUS AFTER A GIVEN AMOUNT OF INBREEDING

Elementary inbreeding theory gives the expected proportion in which that part of the germ plasm initially heterogeneous will be reduced by a given procedure of inbreeding. The progress toward homogeneity is, however, affected by chance at each stage, with the consequence that the amount of heterogenic material varies greatly from one individual to another, though obtained by the same procedure.

More exact knowledge of the nature and extent of this variation is obtainable by two paths: 1. by the calculation of the numbers of junctions formed by recombination at different stages of the inbreeding process, and 2. by the calculation of the rate at which the germ plasm at two different loci ceases to be simultaneously heterogeneous Using these two methods together the variance may be calculated, and the parts ascribed to variation in the lengths of heterogeneous tracts, and in their numbers can be distinguished.

****

D W Goodall (University College of the Gold Coast, Achimota, Gold Coast)

FACTOR ANALYSIS IN PLANT SOCIOLOGY

The methods of factor analysis developed first for the purposes of psychology, and subsequently used in a wide variety of fields, are also suitable for studying the joint distribution in the field of different species of plants, some measure of the quantity of each species of plant present in sample areas providing one variable for the analysis. The results of a set of observations on an area of desert scrub in south-eastern Australia are analysed for purposes of illustration by methods based on those of Hotelling, data for fourteen species being used. Some difficulties likely to arise in the application of factor analysis to problems of plant sociology are discussed.

****

H.C. Hamaker (Philips Research Labs. Eindhoven)

EXPERIMENTAL DESIGN IN INDUSTRY

1 - Once properly understood "experimental design" in the statistical sense does mean a minor revolution in our concepts of technological experimentation.

2 - That only a very small fraction of this revolution has so far been realized is due to the fact that the design of experiment is usually presented via the analysis of variance, a method of presentation exclusively directed towards the mathematical interpretation of the data. Thereby the technological meaning of the analysis is largely lost.

3 - It is possible to present the analysis in a very simple way from which the connection of the various components with the corresponding technological influences can easily be grasped. Such an analysis makes sense even without applying test of significance and probability theory. These statistical techniques should only be brought into play as a final check but should not be seen as the principal aim. Statistical jargon should be carefully avoided. The analysis should be represented in terms of averages and standard-deviations, instead of sums of rows and colums and mean squares.

Whenever possible the result of an analysis should be presented in graphical form. A graph is much more easily understood and remembered than a mean square with double asterisk.

4 - The only way in which the technique can be mastered is by using it. Hence design of experiment should be taught by showing numerous examples of one design and by demonstrating how technological conclusions can be drawn from the analysis. Most textbooks give only one example and then expatiate largely on the mathematical aspects.

5 - The common use of the terms "interaction" and "residue" is confusing and there is a lack of precise definition. We analyse the data into components of the zeroth, first second etc. order, while each component may in its turn be composed of (1) systematic effects and (2) random fluctuations. Experimentally these can be separated by repetition of the experiment.

6 - In agriculture and biology it is usually only possible to carry out one experiment in a year. Hence there is a need for involved designs in order to get the maximum information out of one experiment. Except with life tests, industrial conditions are essentially different in that experiments can be repeated at will. Hence industry does not require too involved designs, but has a

need for wide-scale application of the simpler designs.

7 - Application of statistical techniques to industrial problems is heavily impeded by an exaggerated drive after exactness. If in industry we assume a significance level of 5% it is perfectly satisfactory when the actual level lies say between 3% and 7%. Without statistical techniques people are generally inclined grossly to overestimate the value of their observations and the useful function of statistics is to prevent these gross errors in judgment.

Industrial conditions are perfectly insensitive however against variations in the significance level as indicated above. By purposely disregarding variations of this order we can tremendously simplify statistical techniques and this is of the utmost importance to an effective introduction into industry.

****

M.J **Healy** (Rothamsted Experimental Station)

DECISION BETWEEN TWO ALTERNATIVES, HOW MANY EXPERIMENTS?

When in agriculture or technology a new treatment or process is suggested, the difference in output per unit η between old and new processes can be estimated by the mean $\bar{y}$ of the results of n experiments. The net gain associated with the new treatment may be taken to be a linear function of η, say

$$\eta = k' (\eta - c),$$

where k' is a positive factor depending on the scale of application and c is a constant which depends on the difference in cost of the old and new treatments and the capital cost of the change. The new treatment will be adopted if $\bar{y} - c$ is positive. We wish to arrive at the number of experiments which is economically justifiable allowing for the losses due to possible wrong decisions which in the long run will be added to the cost of experimentation. If, therefore, the cost per unit of experimentation is k and P denotes the probability of getting $\bar{y} - c \geq 0$, we wish to minimize the difference

$$R = kn - P\mu$$
$$= kn - k'P (\eta - c).$$

This expression is the risk function measured from the status quo.

The present note deals with the case analogous to the double sampling procedure of Dodge and Romig in which n has to be decided after a single experiment (or unit of experimentation) has been carried out. For simplicity we assume that the errors in all experiments are independent normal deviates with known standard deviation σ. The probability P is now calculated allowing for the known result y_1 of the first experiment. Moreover, the risk depends on the unknown η which must be eliminated: this has been done by averaging R over the fiducial distribution of η (a normal distribution with mean y_1 and standard deviation σ). Minimisation of the resulting function R provides an intuitively reasonable determination of n, and it has been shown that no other rule can have a uniformly better performance. One notable result brought out by the theory is that it is seldom economic to do a very small amount of additional experimentation. The reason for this is the disproportionate smallness of the chance of altering the decision indicated by the preliminary experimentation. When k, k', σ and $y_1 - c$ are given, the ratios $k\sigma/k$ and $(y_1 - c)/\sigma$ suffice to determine n, and a nomogram for doing this rapidly has been prepared.

The performance of the suggested rule has been evaluated for a series of parameter values. A comparison has been made with the results of an alternative rule in which n is chosen so as to maximize the gain per unit outlay. Some work has also been done on an analogous sequential sampling scheme.

****

T.N. Hoblyn and **S C Pearce** (East Malling Research Station - Kent,

SOME CONSIDERATIONS IN THE DESIGN OF SUCCESSIVE EXPERIMENTS ON FRUIT PLANTATIONS

Some points needing consideration in designing trials with long-lived plants, especially trees, are:

1 - *The initial trial.*

Reliance has usually to be placed on a single trial and consequently it is necessary for the questions under investigation to be posed clearly from the beginning.

Trees are large and there is often only one to a plot. The performance of a single tree is not determined solely by positional effects but is a complex of its history and environment consequently adjustment by covariance on to past performance is often an advantage

2 - *Subsequent trials.*

Plants often outlast the experiment for which they were initially intended and it is then necessary to provide for the application of further treatments when the first set are no longer of interest

Where there is little likelihood of the new and original treatments interacting the new treatments may be applied orthogonally to the blocks or original treatments, or they may be balanced either totally or partially, or they may be supplemented. In this last device one treatment, usually the control, occurs a different number of times from the rest, which are balanced among themselves. From a consideration of available useful design, it is concluded that trials in which a further set of treatments is likely to be called for are best designed, if in randomized blocks, with the number of blocks and of original treatments either equal or differing only by one. In the latter case repeated changes become possible as the residual effects of former treatments disappear.

****

J. W. Hopkins (National Research Council of Canada)

SOME NEEDED TESTS OF SIGNIFICANCE

Requirements for test of significance of the following are illustrated by experimental data. (i) Inconstancy of the negative binomial parameter k characterizing each of mn samples of 2: needed to validate analyses of blood counts before and after imposition of m treatments on groups of n animals assuming that random manipulative and secular discrepancies result in successive counts on the same individual being negatively binomially distributed with a k common to all mn individuals. (ii) Inconstancy of the analogous hypergeometric variance parameter in m samples needed to validate simple criteria for 2-stage acceptance sampling by attributes of large consignments of packaged items (e.g. boxed fruit) when defectives are contagiously distributed between packages. (iii) Departures from goodness of fit of linear regression formulae when both variables are subject to error needed to demonstrate inconsistencies in measurements by two procedures e.g. standard and accelerated methods for moisture in grain. (iv) Inequality of means of m binomial variates of Poisson: needed to demonstrate differences in mean acuity of m groups of subjects in a taste experiment.

****

J. Ipsen (Institute of Laboratories - Boston - Mass.)

FACTORS OF DOSAGE AND HOST- DETERMINING ANTIBODY RESPONSE TO SECONDARY ANTIGEN STIMULUS

A secondary stimulus is an injection of antigen in individuals who have previously been exposed to the same antigen.

The antibody concentration in the serum after a secondary stimulus is dependent on

(1) The immunity status when the secondary stimulus is given.

(2) The antigenic potency of the secondary stimulus.

(3) A negative interaction between the factors of (1) and (2).

The immunity status is in practice estimated in two ways. One is the antibody concentration which can only be used as a comparable function of the immunity status of different individuals if comparable time intervals have elapsed since primary exposure has occurred and if the antibody concentration is above the measurable level in the majority of the individuals. The second estimate can be obtained if the potency of the primary dose is known and the individuals have comparable immunizability. Immunizability is an inherent host characteristic which determines the response to primary stimulus. It can be measured by the dose of antigen which is necessary to confer a given primary immune status.

Immunization of 128 inmates of a school for mentally retarded individuals was performed with two injections of tetanus toxoid 28 days apart. Eight different doses were given in each injection according to a latin square design for the first and second injections. Antibody titers were only measurable in 50 individuals prior to the second injection. The antitoxin titers 14 days after the second stimulus was fitted to the following expectancy formula:

$$Y = a + b_1 x_1 + b_2 x_2 - b_3 x_1 x_2$$

Where Y is log antitoxin titer and x_1 and x_2 is the log potency of the first and second dose, respectively.

A satisfactory fit could only be obtained if the individuals were divided into three groups according to certain somatic criteria, and a parameter c was introduced being constant for each group

$$Y = a + b_1 (x_1 + c) + b_2 x_2 - b_3 x_2 (x_1 + c)$$

The variable c is interpreted as the primary immunizability inherent with the somatic characteristic of the individual.

****

G. **Karreman** (Committee on Mathematical Biology - University of Chicago)

THE MATHEMATICAL BIOLOGY OF THRESHOLD AND RELATED PHENOMENA IN EXCITATION

The existence of a threshold is proved for a physico-chemical model of membrane permeability. The membrane is supposed to consist of one or two molecular layers of a calcium compound (e g. proteinate or lipoproteinate) which is in equilibrium with its ionization products (including calcium). In an electrical field the calcium ions are removed from the site of chemical action and as a result there is a change in the equilibrium concentrations of the calcium compound and its ionization products. The electrical potential across the layer(s) is supposed to be the superposition of the diffusion potential of potassium and the externally applied potential. A treatment is given of the diffusion of the potassium through the membrane the permeability of which to potassium is determined by the equilibrium state of the above mentioned reactions It is shown that the system possesses an unstable equilibrium. From this right orders of magnitude are derived for the chemical and electrical thresholds the increase in permeability upon excitation and the action potential. Excitability curves are derived from the model and shown to be in good agreement with experimental evidence.

Several predictions are made suggesting new experiments From a slight modification of the model repetitive discharges are obtained. The order of magnitude of the potential changes derived from them is right as well as that of their duration.

****

E.A.G. Knowles (Dept. of Engineering Production,
University of Birmingham)

EXPERIMENTAL DESIGNS IN INDUSTRY

With particular reference to production investigations

1 - *Pilot experiments and their interpretation*

Consideration of a variety of different possible interpretations and representations as an aid to obtaining the best guide for the planning of the final investigation appears desirable. The variety of interpretations would often be greater than that which might be thought sufficient from the point of view of the pure statistician because the purpose of the pilot experiment is not only that of giving the best consideration to the particular objectively known scientific and technical conditions of the materials and processes but in addition that of enlisting the co operation of all human personalities involved in every aspect of the work, together with their knowledge and experience.

2 - *Final investigations, their form and interpretation.*

At this stage it is desirable to have the mode of interpretation and the methods of representation agreed beforehand and strictly adhered to, as is usually recommended so that the validity of the statistical significance tests is not threatened by preferred choices. However, renewed experiments with analysis and representation become desirable as soon as the results of the investigation lead to the consideration of further investigations.

3 - *Practical illustration from an investigation connected with tool manufacture.*

The above considerations will be illustrated by means of results obtained in a recent industrial investigation in which the writer has cooperated, the subject being that of hardness variations of standard drills in relation to the various assignable causes given by the raw materials and the methods of production and test.

****

D. C. Lowry (University of California, Berkeley, Cal.)

VARIANCE COMPONENTS WITH REFERENCE TO GENETIC POPULATION PARAMETERS

One of the important problems both of applied and of theoretical genetics is the determination of the comparative effects of the various factors which affect the inheritance of quantitative characters. These characters may be controlled by the action of a large number of gene pairs, by environment and possibly by the interaction of genotype and environment. In analyzing their inheritance Fisher Wright and others have found the analysis of variance a powerful tool in identifying these effects and in characterizing the multigenic system according to additive genetic effects, dominance deviations from the additive scheme and non-allelic gene interaction. Their analyses have been based on models of Mendelian heredity which involve some restrictive assumptions.

The detection and interpretation of components of variance in the study of quantitative traits involve statistical problems of two kinds first, the construction of models and of experimental designs based upon these models, containing fewer restrictive assumptions and, second, the consideration of the purely statistical aspects of the functions of components of variance used as estimates of population parameters.

This paper is intended as a review of what has been accomplished in these two phases of study, particularly the latter

****

L Martin (Faculté de Médecine - Université de Bruxelles).

SUGGESTIONS POUR LA COLLECTION ET L'ANALYSE DE DONNÉES LONGITUDINALES EN GERONTOLOGIE

Supposons qu'un médecin trouve que le cholestérol est normal (1,8 gr/l) chez un patient A et trop élevé (2,8 gr/l.) chez patient B qui manifeste des signes cliniques d'artérioclérose. Ces deux patients sont de même âge, 60 ans par exemple. On peut supposer qu'à 40 ans le cholestérol sanguin était normal dans les 2 cas. Si oui, à partir de quel âge le patient B est-il devenu "hors contrôle"? Si l'on peut déterminer cet instant, la méthode de la médecine préventive aura une bonne chance de réussir, car les lésions au sens large du mot sont peut-être encore réversibles. On propose de déterminer les limites fiduciaires pour la moyenne et une observation isolee à un âge donné en ajustant des polynomes orthogonaux individuels à des données recueillies sur un échantillon de patients suivis d'année en année dès l'âge de 40 ans. L'importance de cet échantillon ainsi que son mode de prélèvement dépendront des facilités techniques. Une telle méthode a été discutée avec le Prof. W.G. Cochran et appliquée dans le cas du développement de l'activité histaminolytique chez 18 femmes suivies pendant les 9 mois de la grossesse. Référence est faite à une suggestion de Sjögen et une autre de Tanner. Ce dernier étudie dans le cas de la croissance des enfants, les mérites respectifs de l'information, tirée de données tranversales (à temps fixe) ou longitudinales (même individu suivi à des moments successifs).
L'auteur fait une suggestion de collection de données a une échelle collaborative intra-et inter pays civilisés.

****

lines of Mendelian genetics. Finally we shall require a theory of variability interpretable in terms of mendelian theory but differing in structure from it, and the recognition of effective units of inheritance whose relation to the genes of Mendelian genetics will require close consideration.

****

E. **Morice** (Institut National de la Statistique et des Etudes Economiques, Paris)

A STATISTICAL STUDY OF THE PHENOMENONS OF HUMAN GROWTH (Exhibit)

The inquiry held in the United States in 1936 contains a great number of documents on the measurements of the numerous elements of the human body in children - boys and girls from 4 to 17 years of age - as well as on the correlations between these measurements at various ages.

The comparison of these various measurements with the height for instance, enables us to determine the relations in allometry (relative growth) of a general form $y = K x^{\alpha}$. One notices that the hypothesis α = constant is confirmed only exceptionally for some dimensions. In general from 4 to 16:

a) for vertical dimensions, α, at first superior to 1 diminishes to become approximately 1 or inferior to 1
b) for horizontal dimensions and perimeters, as well as for the weight, the variations of α are, in general, in the inverse order of the preceding ones and more accentuated.

On the other hand, the correlations between the various measurements - studied for seven of them - show important variations with the age, variations of different types

a) according to the sex
b) according to the nature of the variables placed in correlation.

****

J. Neyman (University of California)

FRISCH'S PROBLEM ON LINEAR STRUCTURAL RELATIONS

The original problem formulated by Ragnar Frisch in 1936 is concerned with two not directly observable random variables ξ and $\eta = \alpha + \beta\xi$ where α and β are unknown constants. Instead of observing the values of ξ and η, we observe those of $X = \xi + U$ and of $Y = \eta + V$ where U and V are random errors of measurement assumed independent of ξ and η and independent of each other. Frisch's question was what must be the distribution of ξ, U and V so that, irrespective of the values of α and β, the regression of Y on X and of X on Y be linear. The purpose of the present paper is to report on the results related to this problem that were obtained by members of the staff of the Statistical Laboratory, University of California, Evelyn Fix, T. Ferguson, Terry A. Jeeves, E.L. Scott and the present author.

****

V. G. Panse (Indian Council of Agricultural Research -New Delhi-India)

PRINCIPLES OF THE SURVEY METHOD OF EXPERIMENTATION

The urgent need of increasing agricultural production by passing on to farmers the results of agricultural research has led to a rapidly growing emphasis in India during the last few years of what may be termed the survey method of experimentation, as against the classical type of experiments at agricultural experiment stations The results from the latter, valuable as they are, cannot be recommended directly for large scale use under actual farming conditions, because the number of experiment stations is small and cannot be regarded as fully representative of the tract served by them. Experiments on a representative sample of the cultivated area therefore become necessary, before recommending a technique to farmer. Such a sample can be secured only by selecting field for experiments randomly out of fields on which a particular crop is grown in the tract

The aim of the experiments is generally to estimate with a reasonable degree of precision the average response to treatments over the tract, detect any interaction of these responses with variation among agriculturally homogeneous sub-division into which tract may be divided and estimate the responses in the individual sub-divisions The precision of the estimates is based on the random variation among fields selected for experiments or rather on the interaction of treatment responses with this variation. Consequently, the importance of the experimental error, as calculated in the classical replicated experiment recedes to the background and replication in the same field. which is essential for estimating the latter, may be sacrificed altogether in order to secure more information on the variation among fields Replication in the sense of the number of repetitions of the experiment in different fields is, of course, important, but randomization of treatments in each field has not the same role in securing the unbiased comparison between the treatments as in the classical experiment. It is, however, safe to adhere to it in order to avoid possible biases arising from border and competition effects peculiar to any fixed arrangement of experimental plots in a field The third principle of replicated experiments, namely, local control involving a compact arrangement of plots and attention to size and shape of plots and blocks is also of little importance, since the variation between fields is far greater than variation between plots in a field which local control is intended to minimise. This allows the latitude needed for fitting the experiment in the farmer's schedule of operations with the least possible disturbance of the latter, which is an essential organisational consideration.

Two sets of projects on the survey type of experiment have been undertaken in India recently In one scheme following the recommendations made by Dr. A.B. Stewart in his report on Soil Fertility Investigations in India, two or three treatments with the local practice as control are tried in each field selected randomly in the tract. The treatments are those which State Departments of Agriculture consider as likely to increase the yield on the basis of the past results at experiment stations. A more ambitious programme of experiments has been commenced this season in several areas under the fertilizers research project sponsored jointly by the Government of India and the T.C.A., with the object of acquiring information on the response of certain important food crops to different nitrogeneous and phosphatic fertilizers

A somewhat different type of experiment is represented by the assessment surveys, conducted by the Indian Council of Agricultural Research in different States in order to estimate the additional yield resulting from various Grow More Food schemes such as distribution of improved seed and fertilizers, provision of new sources of irrigation, tractor ploughing weed infested land etc. Comparison is made between the yield in the area which has received one or more of these aids and an appropriate control. The experiment is not a strictly controlled one in that the treated areas are not selected randomly vis-à-vis the control and the comparison may therefore be open to bias. These surveys form an example of operational research in agriculture and have yielded much valuable data.

****

A. F. Parker-Rhodes - M.A., Ph.D.

ESTIMATING POPULATIONS OF IRREGULARLY OBSERVABLE ORGANISM

The problem treated is that of estimating the total population of an organism in a given area when we cannot assume that all members of the population are ever observable simultaneously. There are many examples of such cases; my own work concerns the higher Basidiomycetes, whose mycelia do not fructify every year and cannot be identified unless they do so.

It is assumed that a permanently and constant population exists in the given area; cases where the population changes appreciably during the period of observation, and where the lifetime of an individual is comparable to the phenological periodicity, are excluded.

The principle of estimation is this. If we have a number N of consecutively numbered counters, known to be less than $\tilde{N}$ but otherwise indeterminate, and sample i of them t times with replacement, the highest ordinal borne by any of the counters taken, $\tilde{n}$, will have a determinate probability distribution depending on t, N, and $\tilde{N}$. Thus, given t, $\tilde{n}$, and $\tilde{N}$, we can estimate N and obtain fiducial limits. The formulae are:

$$E\,[N] = \frac{1-t}{2-t}\,\tilde{n}\left\{\frac{(\tilde{n}/\tilde{N})^{t-2}-1}{(\tilde{n}/\tilde{N})^{t-1}-1}\right\}$$

$$F_{-5\%}\,[N] = \tilde{n}\left\{\left(\frac{\tilde{n}}{\tilde{N}}\right)^{t-1}\frac{1}{20}+\frac{19}{20}\right\}^{\frac{1}{1-t}}$$

$$F_{+5\%}\,[N] = \tilde{n}\left\{\left(\frac{\tilde{n}}{\tilde{N}}\right)^{t-1}\frac{19}{20}+\frac{1}{20}\right\}^{\frac{1}{1-t}}$$

If the frequency distribution of the fraction of the population observable over all the occasions when observations are made should be rectangular, the problem of estimating the total population is formally identical with the above, provided we know the upper limit N. But if the actual frequency distribution is f(x), it can be shown that $n(x) = \int f(x)dx$ is a rectangular variate over the same field, provided that $\operatorname{Lt}_{x\to\infty} n(x)$ is finite; if this condition fails there is no rectangular variate.

In general the form of f cannot be foreseen; in the case of my data on basidiomycetes it can be shown to be at least approximately

true that

$$f(x) = x^{\alpha-1} e^{-x} \cdot \gamma$$

In general, one of the limitations of the method will be that there may be no way of discovering f. On integration, this value of f gives an incomplete gamma function

$$n(x) = \alpha! \left[\left(\log \frac{1}{\beta} \right) x \right] \qquad \lim_{x \to \infty} n(x) = \alpha!$$

Of the two parameters, β can be expressed in terms of α and the (duly weighted) mean of the observed numbers of fructifications $\bar{x}$. The parameter α requires great labour for its rigorous estimation, but for the purposes of this method, which is necessarily rather crude, graphical methods are sufficient.

For each series of observations to be analysed, we first estimate α, and thence compute a series of n values corresponding to the raw xs. From the highest of these we can calculate by the formulae given figures for $F = [N]$, $E[N]$, and $F = [N]$. We can then employ the inverse function to n in order to estimate from these the expectation and fiducial limits of the unknown population X.

These estimates are of course not strictly correct, since in general $f(E[N]) \neq E[f(N)]$, but the error involved is relatively trivial. The expectation distribution is extremely skew, so that upper fiducial limits are often meaningless and represent impossibly large populations. But the estimates of population which the method gives can be generally be relied upon to within 50% either way. This may not sound very good, but most of the error is inherent in the nature of the problem, and more rigorous methods would not usefully reduce it.

Particular examples of the kind of fiducial limits one gets may be cited from my work on Skokholm Island, Wales, which I estimate to have the following total populations of certain basidiomycetes:

Species	lower fiducial limit 5%	Estimate	upper fiducial limit 5%
Naucoria nana	5,000	7,000	50,000
Panaeolus papilionaceus	1,000	2,000	10,000
Clitocybe fragrans	500	1,400	17,000
Panaeolus campanulatus	150	250	900

(these fiducial limits include also errors of sampling in the raw data).

*

S C. Pearce and G.H Freeman (East Malling Research Station, East Malling, Maidstone. Kent)

REPEATED CHANGING OF TREATMENTS IN TRIALS WITH LONG LIVED SPECIES
(Exhibit)

With long-lived species it is a good plan to arrange that a fresh set of treatments can be imposed on the experiment after it has served its original purpose. It is a useful extension to provide for a third set of treatments. The design shown here makes this possible when there are no longer any residual effects of the first set, and indeed permits the introduction of an $(n + 2)$th set when there are no residual effects remaining from any of the first n sets, thus permitting indefinite use of the original experimental material

The top diagram represents a design with seven treatments the two below it show the addition of a second set, either of four or of seven treatments. In the bottom diagrams it is assumed that the original treatments no longer have any residual effects, and a third set, either of four or of seven treatments, is added to each of the middle diagrams

This property is possessed by many experiments designed initially, like this one in balanced incomplete blocks At any stage, an analysis of variance on the data can be carried out readily by solution of the parametric equations.

****

S. Peto (Microbiological Research Department Ministry of Supply
Porton, Wilts)

A DOSE-RESPONSE EQUATION FOR THE INVASION OF MICRO-ORGANISMS

If several batches of test animals are exposed to invading micro-organisms, linearity between log (proportion surviving) and dose is theoretically established, assuming that

(i) the test animals are homogeneous

(ii) the probability of one organism killing its host is constant and small

(iii) the organisms act independently of each other.

The relevant maximun likelihood solution is given and tables recommended to help computation.

When doses are expressed in terms of the ED_{50} a fixed regression line emerges from the underlying hypothesis. Illustrative examples taken from practice are quoted. Comparison with probit analysis is discussed The problem of economical use of test animals is treated The method can be applied to dilution series Bacteriological Research in general might be stimulated and immunological phenomena in particular elucidated if the mode of action of pathogenic organisms is explained by the hypothesis expressed in this paper Lantern slides are used to illustrate essential points.

****

R. **Prigge** (Paul-Ehrlich Institut, Frankfurt a.M.)

DIE ANWENDUNG DER "MUTUNGSBEREICHE" IN DER IMMUNITÄTSFORSCHUNG

Im I. und II. Abschnitt wird die Frage behandelt, welche Rueckschlusse aus den an einer *Stichprobe* gewonnenen Ergebnissen eines Immunisierungsversuches auf die Zusammensetzung der Gesamtpopulation gezogen werden durfen der die Stichprobe entnommen ist. Die verschiedenen Verfahren werden besprochen, um den *"Mutungsbereich"* zu ermitteln, innerhalb dessen der unbekannte wahre Anteil der immunisierten Tiere erwartet werden darf. Es werden Angaben über die Wahrscheinlichkeit der so gewonnenen Aussagen gemacht. Im III. Abschnitt wird ein *Näherungsverfahren* erörtert, dessen Resultate mit den nach den exakten Methoden gewonnenen Ergebnissen gut übereinstimmen und das die Rechenarbeit erheblich reduziert.

Der IV. Abschnitt diskutiert die treffertheoretische und die variationsstatistische Deutung von *Immunisierungskurven* und zeigt die Moglichkeit, mit Hilfe der Mutungsbereiche zu einer Entscheidung zwischen beiden Erklärungsversuchen zu kommen.

Im V. Abschnitt wird die Verwendung der Mutungsbereiche zur Ermittlung des *Wirksamkeitsverhältnisses zweier Impfstoffe* behandelt. Im VI Abschnitt wird die Anwendbarkeit der Naherungslosung auf das Gebiet der *seltenen Ereignisse* nachgewiesen und ihre Verwendung zur Bestimmung des *Keimgehaltes* von Flüssigkeiten besprochen.

****

C. Radhakrishna Rao (Indian Statistical Institute - Calcutta)

A GENERAL THEORY OF DISCRIMINATION WHEN THE INFORMATION ABOUT ALTERNATIVE POPULATION DISTRIBUTIONS IS BASED ON SAMPLES

The problem may be stated as follows

Two samples of sizes n_1 and n_2 are available from two populations $P_1(x \mid \theta_1)$ and $P_2(x \mid \theta_2)$ where x stands for all the measurements and θ for all the parameters. An individual with given measurements y has to be assigned as a member of one of the two groups basing the decision on the observed values only, the parameters occurring in the alternative distributions being unknown. In this paper an attempt is being made to lay down a decision rule independent of the unknown parameters.

If the measurements are p in number we have a total of $(n_1+n_2+1)p$ observations which can be represented by a point in a Euclidean space. The decision rule requires the division of the space into two regions R_1 and R_2 such that when the point of observations falls in R_1 the individual is assigned to the first group and otherwise to the second. Whatever may be the set of regions, it should have the property that errors of classification when the alternative populations are different must be smaller than those when the populations are the same. This criterion leads to the restriction that the size of each region should be the *same* whenever the two probability densities P_1 and P_2 are identical irrespective of what the actual values of the common parameters are

We have now to fix the size of the regions R_1 and R_2 when P_1 and P_2 are identical. When the population distributions are identical, the decision may be equivalent to that of tossing an unbiassed coin so that it is reasonable to take each size as 50 percent. The special case of fixing the size at the 5% level leads to a test of the null hypothesis that the individual belongs to the first group at level 5%, the alternative being the second group.

The problem is now to determine such similar divisions R_1 R_2 covering the entire space which have fixed values when the two distributions are identical and for which the errors of classification is a minimum. The problem of minimising the errors reduces to dividing the region common to the surfaces of sufficient statistics as in the general problem of testing composite hypothesis.

Again, in all cases no uniformly best division is possible on the surfaces of sufficient statistics. We may then determine regions for which the errors of classification is least locally, i.e. for small departures from the equality of populations.

The theory is general and can be applied even when the alternative distributions are more than two.

D. R. Read (Research & Development Dept., The Distillers Co. Ltd., Great Burgh - Epsom - Surrey)

THE DESIGN OF CHEMICAL EXPERIMENTS

The main features that are more or less peculiar to chemical (as contrasted with biological) experimentation are briefly set out, and attention is subsequently confined to the following case which often arises: experimental error small and a prior estimate available, no need for local control, sequential procedure and quantitative factors: the object of the investigation being to estimate a set of optimum conditions for a chemical reaction process, e.g., conditions giving maximum yield or minimum impurity.

A short account is given of the methods already developed by Bo & Wilson, which consist essentially in carrying out experiments in a sequence of small groups, estimating first- or second-degree regressions that give approximate *local* fits to the reaction surface, and using the estimated regression from each group as a guide to the most effective disposition of experimental points in the next group. These methods are then illustrated by an example of their use in a typical problem of chemical development.

****

E. L. Scott (University of California)

BIVARIATE CONTAGIOUS DISTRIBUTIONS

The observed distributions of counts of certain larvae in experimental plots exhibit marks of contagion: if one plot contains a larva then somewhere in the vicinity there was a batch of eggs and this implies an increase in the probability of there being more larvae in the same and neighboring plots. This machinery underlies the single variate contagious distributions deduced by Neyman and later generalized by Beall. The present paper gives formulae for the analogous bivariate distribution that may be useful in closer studies of the behavior of larvae.

****

C A B Smith (The Galton Laboratory University College London)

THE CALCULATION OF CORRELATION BETWEEN COUSINS

For any given character it is possible to find the cousin-cousin correlation by plotting all points (x y) in a correlation diagram in the usual way where x y are the measured values in a pair of cousins. However if two sibs have large numbers of children, every child of one will be a cousin of every child of the other. This will produce a very large number of points in the diagram which may swamp the contribution of the other cousin pairs. It is better to use a three-stage analysis of variance, giving sums of squares (i) between individuals within sibships (ii) between sibships within cousinships (iii) between cousinships. By the use of suitable formulae both the sib-sib and cousin-cousin correlations can then be found.

****

J M Tanner (Sherrington School of Physiology St. Thomas s Hospital, London)

SIZE, SHAPE, AND REGIONAL DIFFERENCES IN THE VERTEBRAL COLUMNS OF INBRED STRAINS OF RABBITS STUDIED BY ANALYSIS OF VARIANCE

(Exhibit)

The lengths and breadths of the bodies of vertebrae numbers 2-29 were measured in two of Dr. P.B. Sawin s strains of inbred rabbits at the Jackson laboratory Bar Harbor, Maine. The rabbits were of various ages from 75 to 1440 days, not all being quite full-grown. The problem was to describe the inter-strain differences in the vertebral columns as succintly as possible, as a basis for genetical experiments on body build in the rabbit

Analysis of variance was used, with a three-way classification by vertebral number (28 classes) race (2 classes) and age (4 classes). The data provided considerable replication within each cell however, an average of 8 rabbits having been measured within each age group of each race The mean values for each cell were the figures analysed with an estimate of the variance of these cell means used as the error mean square.

The triple interaction term and the other interactions have then all clear biological interpretations The conclusion from their significance tests is that one strain differs from the other in (i) general size of all vertebrae (ii) shape of the vertebral column as a whole one strain being slenderer, and (iii) shape of vertebrae in particular and circumscribed regions of the column

****

A. Vessereau (Institut de Statistique Université de Paris)

ENSEIGNEMENT DES MÉTHODES STATISTIQUES APPLIQUEES A LA BIOMETRIE

Ce n est qu après la dernière guerre qu une action coordonnée a été entreprise pour diffuser et appliquer les méthodes statistiques dans différents domaines recouverts par le terme général de "Biométrie".

Un enseignement de ces méthodes a été créé en 1946 à l'Institut de Statistique de l Université de Paris, il a été suivi dès le début et continue à être suivi par les étudiants qui se destinent à la recherche agronomique dans les territoires de la France d Outre-Mer ainsi que par des chercheurs déja affectés à des Centres de Recherches. Quelques années plus tard l Institut de Statistique a également créé un enseignement de "Génétique des populations".

Il y a quelques années l'Institut National Agronomique a inscrit l enseignement des méthodes statistiques au programme de sa troisième année d étude (élèves se destinant à des carrières de recherches généticiens pédologues, etc . .) des compleménts de mathématique et en particulier des notions de calcul des probabilités preparatoires à cet enseignement sont donnés au cours des deux premières années d étude.

Dans le cadre de la Faculté des Sciences, les étudiants qui préparent le certificat de "génétique" reçoivent les notions de statistique indispensables en cette matière. Les étudiants en psychologie appliquée (Institut de psychologie de l Université de Paris) reçoivent également un enseignement statistique de base et un enseignement spécialisé.

Si, dans les secteurs "biologie" et "agronomie", l'enseignement des méthodes statistiques se trouve organisé de façon à peu près cohérente par contre, dans le secteur "medical", il n y a eu jusqu'ici que des tentatives fragmentaires (quelques séries de conférences) insuffisamment coordonnées: des progrès importants restent à faire dans ce domaine.

Une des principales difficultés rencontrées dans l'enseignement des méthodes statistiques réside dans l'insuffisance de formation mathématique de la pluparte des élèves (leurs études antérieures ont été orientées surtout vers les sciences naturelles) et dans la nouveauté que présente pour eux le mode de raisonnement probabiliste. Cette situation pourrait être améliorée si une initiation statistique et probabiliste était donnée au cours des études secondaires, et si certaines modifications étaient apportées au programme des connaissances exigées des candidats aux carrières biologiques et médicales.

Dans l enseignement dispensé à des étudiants de formation mathématique parfois précaire il est souvent préférable de renoncer au exposés rigoureux et de faire appel à des raisonnements approchés ou intuitifs sans toutefois masquer les difficultés le strict domaine d'application des méthodes doit d'autre part être bien précisé

Il a été reconnu que, sans sacrifier le notions théoriques essentielles, l enseignement doit rester aussi concret que possible. Les exercices pratiques sont indispensables et peuvent prendre plusieurs formes· établissement ou vérification de lois statistiques et propriétés des échantillons, à partir de tirages de boules jeux de dès ou de cartes petits problèmes destinés a faciliter l assimilation des parties théoriques de l'enseignement applications numériques avec exécution complète des calculs et emploi de machines à calculer.

****

LIST OF MEMBERS (Additions)

Dr. H.C. BATSON	Dept. of Public Health 1800 West Fillmore St. **CHICAGO 12** (U S A.)
Prof. H. BERGSTRÖM	Chalmers University of Technology Dept. of Appl. Mathematics **GÖTEBORG** (Sweden)
Prof. J. BERKSON	Mayo Clinic **ROCHESTER. Minn.** (U.S.A.)
Dr. B.B. DAY	3136 Dumbarton Ave N.W. **WASHINGTON 7 D.C** (U.S.A.)
Prof. M. FRECHET	2, Rue Emilie Faguet, 14^{e} **PARIS** (France)
Dr. H.C. HAMAKER	Philips Research Laboratories **EINDHOVEN** (The Netherlands)
Dr. L.B. HOLT	The Wright Fleming Institute St. Mary's Hospital Paddington, **LONDON W.2** (England)
Prof. G. VAN ITERSON	Eemnesserweg 89 **BEARN** (The Netherlands)
Miss. E.A.G. KNOWLES	26 Westfield Rd. Edgbaston **BIRMINGHAM 15** (England)
Prof. E.J. MACDONALD	2310 Baldwin St. **HOUSTON 6** Texas (U.S.A.)
Prof. J. NEYMAN	Statistical Laboratory University of California **BERKELEY 4** California (U.S.A.)
Dr. E. OLBRICH	Müllerstr. 59 **INNSBRUCK** (Austria)

Dr. V. G. PANSE — Indian Council of Agricultural Research, NEW DELHI (India)

Dr. H. PRÖBSTEL — Eberstädterstr. 13, PFUNGSTADT/HESSEN (Germany)

Dr. W. RODDEWIG — Walluferstrasse 4, ELTVILLE/RHEIN (Germany)

Dr. E. L. SCOTT — Statistical Laboratory, University of California, BERKELEY 4 California (U.S.A.)

****

Approved For Release 2004/02/11 : CIA-RDP80-00809A000500400179-7

J. Berkson (Division of Biometry and Medical Statistics, Mayo Clinic, Rochester, Minnesota, U.S.A.)

MAXIMUM LIKELIHOOD AND MINIMUM CHI-SQUARE ESTIMATES OF REGRESSION COEFFICIENTS

Several functions will be briefly discussed, but chiefly attention will be given to the logistic function, $P = 1/1+e^{-(\alpha + \beta x)}$ with the observation on P assumed to be a random binomially distributed variable, as in the model for bioassay with quantal response. Three estimators are investigated: (1) Maximum likelihood, (2) Minimum χ^2 (Pearson), and (3) Minimum logit χ^2 (Berkson, J., J. Am. Statist. A., 39 [1944] 357).

Two cases are considered, (1) β known, α to be estimated and (2) α and β both to be estimated. In the example dealt with, there are three equally spaced values of x ("doses" in bioassay) with N = 10 at each. The dose arrangements are for P = 0.3, 0.5, 0.7, corresponding respectively to the three consecutive doses, and for other sets of three doses each, in which the value of P corresponding to the central dose is 0.6, 0.7, 0.8, 0.85.

In the case with β known, α to be estimated, the results are based on calculations of the total sampling population; in the case with both α and β to be estimated, they are based on a sample of 2,000 with each dose arrangement.

For central dosage P = 0.5 each of the three estimators is unbiased; for other dosage arrangements each is biased, the maximum likelihood estimate positively, each of the χ^2 estimates negatively. The mean square error and variance about the mean are largest for the maximum likelihood estimate, smaller for the minimum Pearson χ^2 estimate, and smallest for the minimum logit χ^2 estimate.

The estimates are considered in relation to the Cramer-Rao lower bound for the mean square error. The bound value itself is highest for the maximum likelihood estimate, lower for the minimum χ^2 estimate, and lowest for the minimum logit χ^2 estimate, but the m.s.e. of all three estimates is higher than their respective lower bound value.

Each of the estimators is sufficient. Blackwell's theorem (Ann. Math. Statis. 18 [1947] 105), may therefore appropriately be applied. The "Blackellized" value of the estimate (conditional expectation of estimate for fixed value of sufficient statistic) is the same for the maximum likelihood estimate as before Blackwellization, but with the minimum logit χ^2 estimate, the mean square error is diminished by Blackwellization to its lower bound value.

****

L. B. Holt (Wright Fleming Institute. London)

QUANTITATIVE STUDIES IN DIPHTERIA PROPHYLAXIS

An attempt has been made to characterise the antigenicity of any diphtheria prophylactic in mathematical terms (responses to a single inoculation).

Use is made of the observation that the responses among a group of similar subjects, identically treated, is log. normally distributed; as well as the observation that when the results of a dose-response experiment are plotted as probit y% against log. dose administered a straight line is obtained - the probit regression line.

It was found that three variables are involved, namely d. the dose required to produce some arbitrary reference point of response, b., the slope of the probit regression line, which gives information in respect of the percentage increase of subjects attaining or exceeding some arbitrary level of response with increase of dose, and $\sigma_{log.}$, the standard deviation of logs of titres. These three variables are incorporated into one general equation:

$$\text{Log. G.M.}_{(\text{dose D.})} = \text{log. G.M.}_{(\text{dose d.})} + \sigma_{\text{logs.}}\, b.\, (\text{log.} \frac{\text{dose D.}}{\text{dose d.}})$$

The product of $\sigma_{\text{logs.}}$ and b. for any one set of data is a constant (K.) which is numerically equal to the slope of the dose-response curve log. G.M./log. dose.

The value of K. for the diphtheria prophylactic P.T.A.P. measured in children was found to be approximately 0.86 and in guinea pigs about 0.6; and the $\sigma_{log.}$ of titres 0.65 for children and 0.51 for guinea pigs.

Evidence is offered to show a marked dissimilarity in the constants for other prophylactics measured in children and in guinea pigs; sometimes the guinea pig will underrate a prophylactic in terms of children, and overrate it for another kind of prophylactic.

The importance of the laboratory use of appropriate "Standard Antigens" that have been calibrated in the field is stressed.

****

M. Keuls (Institute of Horticultural Plant Breeding, Wageningen)

TESTING DIFFERENCES IN AN ANALYSIS OF VARIANCE

The problem: in applying an analysis of variance, after having concluded from an F-test at a significance level 0,05 that the null-hypothesis $\mu_1 = \mu_2 = \mu_m$ has to be rejected, so that at least some of the μ_i are different, most research workers feel the lack of a convenient statistical procedure stating *what differences* should be considered real. In general we may be interested in the question what contrasts within a previously chosen subset of the set of all contrasts $\sum a_i \mu_i$ ($\Sigma a_i = o$) should be considered different from null.

A procedure in general use is the least significant difference test (l.s.d.-test), consisting of applying an ordinary *t*-test to each difference or contrast seperately, then and only then, if the F-test rejects the null-hypothesis. This procedure has been denounced by most writers to-day, as it highly exaggerates the significance of the conclusions.

Several alternative procedures which fall into two classes have been suggested. The oldest are the multilayer significance tests (Newman, Duncan, Tukey, Keuls). Although these tests implicate some nominal significance level, the real significance levels are problematic. The other class contains procedures of multiple confidence statements (Tukey, Scheffe, Roy, Bose and Roy and others). Here nominal and real significance levels are identical. The procedures are simple and will be clear also to non-statistically trained readers of the records.

In discussing old and new procedures, the following points are of interest:

1. There may be defined different significance levels or according to prof. Tukey "error rates".
2. There may be a choice between a significance level procedure and a procedure of confidence statements.
3. A point that has been insufficiently stressed by writer on confidence procedures, is to indicate, before choosing a procedure,

A.R.G. Owen, (Department of Genetics, Cambridge).

EXPERIMENTAL DESIGN IN GENETICS

Analytical genetics is a suitable term to describe that kind of genetical research where, by controlled breeding of organisms, we follow the assortment and recombination of sets of discrete Mendelian factors. The analysis and interpretation of the data which arise involve a coherent system of statistical methods, exhibiting points both of analogy and contrast with those so well-known in the realm of agricultural experimentation.

In analytical genetics we are concerned fundamentally with the estimation of certain pure numbers such as segregation ratios or recombination fractions; i.e. those parameters which enable the breeding behaviour of organisms to be predicted. In agriculture we tend to stress the detection of differences between stocks of treatments. This however is clearly a question of estimation and the difference is only one of emphasis.

In agricultural statistics almost everything involves normally distributed variates, and the prime tool is the analysis of variance, with the principle of orthogonality as guiding consideration allowing individual effects to be distinguished. In genetics the variates are always whole numbers being multinomial class frequencies. Taking estimation as the basic problem, complete coherence of method is acheived by proceeding from the maximum likelihood theory, using the technique of scores, a mode of presentation whose power is not yet always fully appreciated.

Scores are linear functions of the class frequencies, and they lead directly to the analysis of Chi-Squared which is our prime tool and the counterpart of the analysis of variance. The analysis of Chi-squared can be arranged to exhibit such features as orthogonality, component effects, interaction and error terms.

Both discriminant functions for grouping and the normal theory of curvilinear regression have their Chi-squared analogues in this field.

The Latin Square occurs inevitably and essentially in the design of multiple point linkage tests, when we wish to separate Mendelian ratios from viability effects, but it is used in a way peculiar to the subject. For instance the feature of randomisation is absent.

****

M. Strecker.

DESIGN AND IMPLEMENTATION OF A SAMPLE ON MILK PRODUCTION IN AGRICULTURAL AND FOREST FARMS OF BAVARIA IN THE AGRICULTURAL YEAR 1951/1952

In the compass of a research commission the Bavarian Statistical Office implemented a sample on milk production in the agricultural and forest farms of Bavaria from 0.5 ha of agricultural area and more for the agricultural year 1951/52.

The inquiry was designed as a stratified sample with "cow-keeping agricultural farm" as unit of selection. For an efficient stratification the following three main strata were considered:

a) farms from 0 to less than 50 ha of agricultural area (sampling ratio 2 per cent)
b) farms from 50 to less than 150 ha of agricultural area (sampling ratio 10 percent)
c) farms comprising 150 and more ha of agricultural area (complete enumeration).

On the whole, about 9, 100 cow-keeping agricultural farms in Bavaria (2.15 per cent of the total) were covered.

The cow keepers selected had to report the cow's milk produced on their farms on a fixed day every month on a report form. The collaboration of the cow-keeping farms included at random in the sample was voluntary. Therefore only about half the cow keepers selected had filled in their report forms. As is well known, the disregard of such cases may bias the estimates of the sample to an unknown extent. For this reason detailed investigations on the problem of non-response as defined by K. Hansen and W. Hurwutz were made: the population of the respondents may be divided into two groups:

Group A = those who are readu to answer and would send back the completed report form;
Group B = all others not inclined to answer.

In selecting the respondents, the contingency must be taken into account that a respondent can be found who responds. In a complete specification of this item every respondant would have to be given a probability measuring this contingency.

As estimate for the milk production x'

$x' = \frac{N}{n} (m \bar{x}_1 + s \bar{x}_2'')$ may be accepted. The mean deviation belonging to it is

$$\sigma_{x'} = N^2 \frac{N - n}{(n - 1)n} \sigma^2 + \frac{N}{n} \frac{S^2}{S-1} (k - 1) \sigma_b^2$$

Here are:

N = number of cow-keeping farms

n = number of farms selected

m = number of answers received

s = n - m = number of non-respondents

S = total number of cow-keeping farms unwilling to answer

r = number of cow keepers selected from the s cow keepers for another requiry

$k = \frac{s}{r}$

σ^2 = dispersion of the N cow keepers

σ_b^2 = dispersion of non-respondent cow keepers

$\bar{x}'_1$ = average milk production from the m answering cow keepers

$\bar{x}''_2$ = average milk production from the r cow keepers covered in addition

The result was that the sample was not biased by non-response. The mean square deviation of the NON-RESPONSE problem was for $\bar{x}'_1$, 0.8 per cent, the mean square deviation for $\bar{x}'$, without taking into account the non-responding cow keepers, 1.0 per cent.

****

Timing

compl. fertilizer

NOTES

$$y = y_0 + d(1 - 10^{-kx})$$ Mitscherlich (?)

$$x = \frac{1}{k} \log \frac{2.302Kd}{M}$$ optim. M = const.

$$y = a + bx + cx^2$$

$$x = \frac{M - b}{2c}$$ optimum.

A DOSE-RESPONSE EQUATION FOR THE INVASION OF MICRO-ORGANISMS

S. Peto

Microbiological Research Department, Ministry of Supply, Porton, Wilts.

The Problem

A series of k doses (n_1 , n_2 , $\cdots$, n_k micro-organisms) is administered to m_1 , m_2 , $\cdots$, m_k test animals respectively, and the corresponding survivors are observed to number r_1 , r_2 , $\cdots$, r_k . (The terms "survivors" and "killed" whenever they occur subsequently, are meant to cover also the case when "not infected" are compared with "infected"). This paper derives a dose-response relation from a hypothesis based on the mode of action of the micro-organisms against their host first suggested by H. A. Druett (2).

Assuming that

(i) the test animals are homogeneous,
(ii) the probability of one organism killing its host is p (small),
(iii) the organisms act independently of each other,

then if n is the number of organisms administered to each subject, the expected proportion surviving is given by

$$S = (1 - p)^n \doteq e^{-np} \tag{1}$$

The statistical problem is to estimate the single parameter p from a series of observations of corresponding values of the actual proportions surviving r/m and the numbers of attacking organisms n. Equation (1) may be written

$$\ln S = -pn \tag{1a};$$

thus if we plot

$$\ln \frac{r_1}{m_1}, \ln \frac{r_2}{m_2}, \cdots, \ln \frac{r_k}{m_k}$$

against n_1 , n_2 , $\cdots$, n_k and fit a straight line to the resulting points, the negative slope of this line will represent p, the probability of any one organism killing an animal.

It follows that dose-response relationships for all kinds of organisms and test animals should be represented by a single line of constant slope when the doses are expressed as multiples of the ED_{50} . For taking c

organisms to be the ED_{50}, equation (1a) becomes

$$\ln 0.5 = -pc \tag{1b}$$

and putting $n = fc$ we obtain

$$\ln S = f \ln 0.5 \tag{2}$$

The Maximum Likelihood Solution

Let the probability that one particular animal survives be

$$P = (1 - p)^n \doteq e^{-np};$$

then the probability of r surviving out of m animals at risk will be

$$Pr\{r \text{ surv. of } m\} = \binom{m}{r} P^r (1 - P)^{m-r} \doteq \binom{m}{r} e^{-npr} (1 - e^{-np})^{m-r}$$

and the logarithm of the likelihood is

$$L = -p \sum n_i r_i + \sum (m_i - r_i) \ln (1 - e^{-n_i p}) + \text{const.}$$

Hence

$$\frac{dL}{dp} = -\sum n_i r_i + \sum (m_i - r_i) n_i \frac{e^{-n_i p}}{1 - e^{-n_i p}}$$

$$= -\sum m_i n_i + \frac{1}{p} \sum (m_i - r_i) \frac{n_i p}{1 - e^{-n_i p}} \tag{3}$$

and

$$\frac{d^2 L}{dp^2} = -\sum (m_i - r_i) n_i^2 \frac{e^{-n_i p}}{(1 - e^{-n_i p})^2}$$

$$= -\frac{1}{p^2} \sum (m_i - r_i)(n_i p)^2 \frac{e^{n_i p}}{(e^{n_i p} - 1)^2} \tag{4}$$

Substituting $n_i p = x_i$ equation (3) takes the following form for L_{max}:

$$f(p) = -\sum m_i n_i + \frac{1}{p} \sum (m_i - r_i) \frac{x_i}{1 - e^{-x_i}} = 0 \tag{5}$$

and using the same substitution for equation (4) the variance of p is given by

$$\text{Var } (p) = \frac{-1}{\frac{d^2 L}{dp^2}} = \frac{p^2}{\sum (m_i - r_i) \frac{x_i^2 e^{x_i}}{(e^{x_i} - 1)^2}} \tag{6}$$

Equation (5) can be readily solved to any degree of accuracy with the help of the Table A and the application of Newton's Method of approxi-

mation; Table B evaluates expression (6) in order to obtain fiducial limits to p.

The goodness of fit can be tested with χ^2, and confidence limits can be assigned at any desired level of dose or response.

A Numerical Example

This department obtained the following data (first 4 columns of Table I) from one of a series of experiments in which *B. anthracis* spores were allowed to enter guinea pigs by the respiratory route (3).

The necessary steps to complete Table I and to obtain the required results are as follows:

(1) After having entered the values r/m in the appropriate column, ln r/m is plotted against n (preferably on log paper) and a line through the origin is fitted by eye; an estimate $p_1(0.019)$ of the slope p is thus obtained (see fig. 1).
(2) The np_1 values are computed and entered in the x-column, leaving room for corresponding values of a second cycle (np_2).
(3) Tables A and B give the following values:

x	$x(1 - e^{-x})$	$x^2e^x/(e^x - 1)^2$
0.32	1.169	0.992
0.66	1.366	0.964
1.23	1.738	0.883
1.91	2.212	0.744

The 4 values $x(1 - e^{-x})$ are multiplied in turn by 8, 18, 21, 28($m - r$ column) the sum of these products being 133.214; the 4 values $x^2e^x/(e^x - 1)^2$ are treated likewise yielding 64.66. Then the residue of equation (5) $f(p_1) = 80$ and Var $(p) = 0.00000558$ are obtained as set out in the lower part of Table I.

(4) A second approximation p_2 to p is computed by applying Newton's Method: $p_2 = p_1 + f(p_1)$ Var $(p) = 0.0191$.
(5) In the x-column the new pn values are entered in brackets (0.33 ··· 1.95), and $f(p_2) = +3$ is evaluated using Table A. Clearly no closer approximation is required and in fact Var (p) need not be recomputed.
(6) The expected numbers of survivors are calculated, and used to evaluate χ^2 with degrees of freedom one less than the number of doses, since one parameter has been estimated from the data.

TABLE A

$x/(1 - e^{-x})$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
.0	1.000	1.005	1.010	1.015	1.020	1.025	1.030	1.035	1.041	1.046
.1	1.051	1.056	1.061	1.066	1.072	1.077	1.082	1.087	1.093	1.098
.2	1.103	1.109	1.114	1.119	1.125	1.130	1.136	1.141	1.147	1.152
.3	1.157	1.163	1.169	1.174	1.180	1.185	1.191	1.196	1.202	1.208
.4	1.213	1.219	1.225	1.230	1.236	1.242	1.248	1.253	1.259	1.265
.5	1.271	1.277	1.282	1.288	1.294	1.300	1.306	1.312	1.318	1.324
.6	1.330	1.336	1.342	1.348	1.354	1.360	1.366	1.372	1.378	1.384
.7	1.391	1.397	1.403	1.409	1.415	1.421	1.428	1.434	1.440	1.446
.8	1.453	1.459	1.465	1.472	1.478	1.485	1.491	1.497	1.504	1.510
.9	1.517	1.523	1.530	1.536	1.543	1.549	1.556	1.562	1.569	1.575
1.0	1.582	1.589	1.595	1.602	1.609	1.615	1.622	1.629	1.635	1.642
1.1	1.649	1.656	1.662	1.669	1.676	1.683	1.690	1.697	1.703	1.710
1.2	1.717	1.724	1.731	1.738	1.745	1.752	1.759	1.766	1.773	1.780
1.3	1.787	1.794	1.801	1.808	1.815	1.822	1.830	1.837	1.844	1.851
1.4	1.858	1.865	1.873	1.880	1.887	1.894	1.902	1.909	1.916	1.924
1.5	1.931	1.938	1.946	1.953	1.960	1.968	1.975	1.982	1.990	1.997
1.6	2.005	2.012	2.020	2.027	2.035	2.042	2.050	2.057	2.065	2.072
1.7	2.080	2.088	2.095	2.103	2.110	2.118	2.126	2.133	2.141	2.149
1.8	2.156	2.164	2.172	2.180	2.187	2.195	2.203	2.211	2.219	2.226
1.9	2.234	2.242	2.250	2.258	2.266	2.273	2.281	2.289	2.297	2.305
2.0	2.313	2.321	2.329	2.337	2.345	2.353	2.361	2.369	2.377	2.385
2.1	2.393	2.401	2.409	2.417	2.425	2.433	2.442	2.450	2.458	2.466
2.2	2.474	2.482	2.490	2.499	2.507	2.515	2.523	2.532	2.540	2.548
2.3	2.556	2.565	2.573	2.581	2.589	2.598	2.606	2.614	2.623	2.631
2.4	2.639	2.648	2.656	2.665	2.673	2.681	2.690	2.698	2.707	2.715
2.5	2.724	2.732	2.741	2.749	2.757	2.766	2.774	2.783	2.792	2.800
2.6	2.809	2.817	2.826	2.834	2.843	2.851	2.860	2.869	2.877	2.886
2.7	2.895	2.903	2.912	2.920	2.929	2.938	2.946	2.955	2.964	2.973
2.8	2.981	2.990	2.999	3.007	3.016	3.025	3.034	3.043	3.051	3.060
2.9	3.069	3.078	3.086	3.095	3.104	3.113	3.122	3.131	3.139	3.148
3.0	3.157	3.166	3.175	3.184	3.193	3.202	3.211	3.219	3.228	3.237
3.1	3.246	3.255	3.264	3.273	3.282	3.291	3.300	3.309	3.318	3.327
3.2	3.336	3.345	3.354	3.363	3.372	3.381	3.390	3.399	3.408	3.417
3.3	3.426	3.435	3.445	3.454	3.463	3.472	3.481	3.490	3.499	3.508
3.4	3.517	3.527	3.536	3.545	3.554	3.563	3.572	3.581	3.591	3.600
3.5	3.609	3.618	3.627	3.637	3.646	3.655	3.664	3.673	3.683	3.692
3.6	3.701	3.710	3.720	3.729	3.738	3.747	3.757	3.766	3.775	3.785
3.7	3.794	3.803	3.812	3.822	3.831	3.840	3.850	3.859	3.868	3.878
3.8	3.887	3.896	3.906	3.915	3.924	3.934	3.943	3.952	3.962	3.971
3.9	3.981	3.990	3.999	4.009	4.018	4.028	4.037	4.046	4.056	4.065
4.0	4.075	4.084	4.093	4.103	4.112	4.122	4.131	4.141	4.150	4.160
4.1	4.169	4.179	4.188	4.198	4.207	4.216	4.226	4.235	4.245	4.254
4.2	4.264	4.273	4.283	4.292	4.302	4.312	4.321	4.331	4.340	4.350
4.3	4.359	4.369	4.378	4.388	4.397	4.407	4.416	4.426	4.436	4.445
4.4	4.455	4.464	4.474	4.483	4.493	4.503	4.512	4.522	4.531	4.541
4.5	4.551	4.560	4.570	4.579	4.589	4.599	4.608	4.618	4.627	4.637
4.6	4.647	4.656	4.666	4.676	4.685	4.695	4.705	4.714	4.724	4.733
4.7	4.743	4.753	4.762	4.772	4.782	4.791	4.801	4.811	4.820	4.830
4.8	4.840	4.850	4.859	4.869	4.879	4.888	4.898	4.908	4.917	4.927
4.9	4.937	4.946	4.956	4.966	4.976	4.985	4.995	5.005	5.014	5.024

Abridged from "A Table of the Function $(e^x/x) = x/(1 - e^{-x})$ and its Applications to Problems in Compound Interest" by J. F. Steffensen, Skandinavisk Aktuarietidskrift, 1938; by kind permission of the author and the publishers.

TABLE B

$x^2e^x/(e^x - 1)^2$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
.1	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.997	0.997
.2	0.997	0.996	0.996	0.996	0.995	0.995	0.994	0.994	0.993	0.993
.3	0.993	0.992	0.992	0.991	0.990	0.990	0.989	0.989	0.988	0.987
.4	0.987	0.986	0.985	0.985	0.984	0.983	0.983	0.982	0.981	0.980
.5	0.979	0.979	0.978	0.977	0.976	0.975	0.974	0.973	0.972	0.971
.6	0.971	0.970	0.969	0.968	0.967	0.966	0.964	0.963	0.962	0.961
.7	0.960	0.959	0.958	0.957	0.956	0.954	0.953	0.952	0.951	0.950
.8	0.948	0.947	0.946	0.945	0.943	0.942	0.941	0.939	0.938	0.937
.9	0.935	0.934	0.932	0.931	0.930	0.928	0.927	0.925	0.924	0.922
1.0	0.921	0.919	0.918	0.916	0.915	0.913	0.911	0.910	0.908	0.907
1.1	0.905	0.903	0.902	0.900	0.898	0.897	0.895	0.893	0.892	0.890
1.2	0.888	0.886	0.885	0.883	0.881	0.879	0.878	0.876	0.874	0.872
1.3	0.870	0.868	0.867	0.865	0.863	0.861	0.859	0.857	0.855	0.853
1.4	0.852	0.850	0.848	0.846	0.844	0.842	0.840	0.838	0.836	0.834
1.5	0.832	0.830	0.828	0.826	0.824	0.822	0.820	0.818	0.816	0.814
1.6	0.811	0.809	0.807	0.805	0.803	0.801	0.799	0.797	0.795	0.792
1.7	0.790	0.788	0.786	0.784	0.782	0.780	0.777	0.775	0.773	0.771
1.8	0.769	0.767	0.764	0.762	0.760	0.758	0.755	0.753	0.751	0.749
1.9	0.747	0.744	0.742	0.740	0.738	0.735	0.733	0.731	0.729	0.726
2.0	0.724	0.722	0.720	0.717	0.715	0.713	0.710	0.708	0.706	0.704
2.1	0.701	0.699	0.697	0.694	0.692	0.690	0.687	0.685	0.683	0.681
2.2	0.678	0.676	0.674	0.671	0.669	0.667	0.664	0.662	0.660	0.657
2.3	0.655	0.653	0.651	0.648	0.646	0.644	0.641	0.639	0.637	0.634
2.4	0.632	0.630	0.627	0.625	0.623	0.620	0.618	0.616	0.614	0.611
2.5	0.609	0.607	0.604	0.602	0.600	0.597	0.595	0.593	0.590	0.588
2.6	0.586	0.584	0.581	0.579	0.577	0.574	0.572	0.570	0.568	0.565
2.7	0.563	0.561	0.559	0.556	0.554	0.552	0.549	0.547	0.545	0.543
2.8	0.540	0.538	0.536	0.534	0.532	0.529	0.527	0.525	0.523	0.520
2.9	0.518	0.516	0.514	0.512	0.509	0.507	0.505	0.503	0.501	0.498
3.0	0.496	0.494	0.492	0.490	0.488	0.485	0.483	0.481	0.479	0.477
3.1	0.475	0.473	0.470	0.468	0.466	0.464	0.462	0.460	0.458	0.456
3.2	0.454	0.452	0.449	0.447	0.445	0.443	0.441	0.439	0.437	0.435
3.3	0.433	0.431	0.429	0.427	0.425	0.423	0.421	0.419	0.417	0.415
3.4	0.413	0.411	0.409	0.407	0.405	0.403	0.401	0.399	0.397	0.395
3.5	0.393	0.391	0.389	0.388	0.386	0.384	0.382	0.380	0.378	0.376
3.6	0.374	0.372	0.371	0.369	0.367	0.365	0.363	0.361	0.359	0.358
3.7	0.356	0.354	0.352	0.350	0.349	0.347	0.345	0.343	0.342	0.340
3.8	0.338	0.336	0.334	0.333	0.331	0.329	0.328	0.326	0.324	0.322
3.9	0.321	0.319	0.317	0.316	0.314	0.312	0.311	0.309	0.307	0.306
4.0	0.304	0.302	0.301	0.299	0.298	0.296	0.294	0.293	0.291	0.290
4.1	0.288	0.286	0.285	0.283	0.282	0.280	0.279	0.277	0.276	0.274
4.2	0.273	0.271	0.270	0.268	0.267	0.265	0.264	0.262	0.261	0.259
4.3	0.258	0.256	0.255	0.254	0.252	0.251	0.249	0.248	0.246	0.245
4.4	0.244	0.242	0.241	0.239	0.238	0.237	0.235	0.234	0.233	0.231
4.5	0.230	0.229	0.227	0.226	0.225	0.224	0.222	0.221	0.220	0.218
4.6	0.217	0.216	0.215	0.213	0.212	0.211	0.210	0.208	0.207	0.206
4.7	0.205	0.203	0.202	0.201	0.200	0.199	0.197	0.196	0.195	0.194
4.8	0.193	0.192	0.190	0.189	0.188	0.187	0.186	0.185	0.184	0.183
4.9	0.181	0.180	0.179	0.178	0.177	0.176	0.175	0.174	0.173	0.172

Abridged from "A six place Table of the Einstein Functions" by J. Sherman and R. R. Ewell, The Journal of Physical Chemistry, June 1942 by kind permission of The Williams & Wilkins Company, Baltimore.

TABLE I.

Computations for the Fitting of a Dose – ln Proportion surviving Regression Equation

n	m	r	$m-r$	r/m	x		$S=e^{-x_0 x}$	mS	$r-mS$	$\frac{(r-mS)^2}{mS(1-S)}$
Dose in 100 Spores	No. of animals at risk	Surviving	Killed	Prop. Surv. observed	np_1	(np_2)	Prop. Surv. expected	Survivors expected		
16.8	32	24	8	0.75	0.32	(0.33)	0.72	23.0	+1.0	0.16
31.7	32	14	18	0.44	0.66	(0.67)	0.51	16.3	−2.3	0.66
64.6	32	11	21	0.34	1.23	(1.25)	0.29	9.3	+1.7	0.44
100.5	32	4	28	0.125	1.91	(1.95)	0.14	4.5	−0.5	0.06
216.6	128	53	75							

$x^2 = 1.32$

using equation (5):

$$-\sum mn = -216.6 \times 32 = -6931$$

$$\frac{1}{p_1}\sum (m-r)\frac{x}{1-e^{-x}} = \frac{133.214}{0.019} = 7011$$

$$\frac{f(p_1)}{} = +80$$

$$\ln S = -0.0194n$$

$$\mathrm{Var}\,(p) = \frac{0.000361}{64.66} = 0.0000558 \quad \text{(equation 6)}$$

$$p_2 = p_1 + f(p)\,\mathrm{Var}\,(p) = 0.019 + 80 \times 558 \times 10^{-7} = 0.0194$$

$$p_2' = p_2 + 1.96\sqrt{\mathrm{Var}\,(p)} = 0.0240$$

$$p_2'' = p_2 - 1.96\sqrt{\mathrm{Var}\,(p)} = 0.0148$$

95% fid. Lts.

ED_{50} (using equation 1b)

95% lower Lim: $\ln 0.5/(-p_2')$	$\ln 0.5/(-p_2)$	95% upper Lim: $\ln 0.5/(-p_2'')$
28.9 ⟵	35.7 ⟵	46.8

Experimental Verification of the Hypothesis

The extent to which the basic hypothesis fits the experimental facts is shown by the figures in Table II which gives the pooled values of χ^2 for several experiments on these different micro-organisms. The individual responses are plotted in figs. 2a, b, c; the doses have been expressed as multiples of the ED_{50}'s, and the lines have the theoretical slope of ln 0.5. It should be noted that in the case of *B. anthracis* the experiments were not replicates since the particle size of the cloud was made to vary considerably from experiment to experiment, so that individual regres-

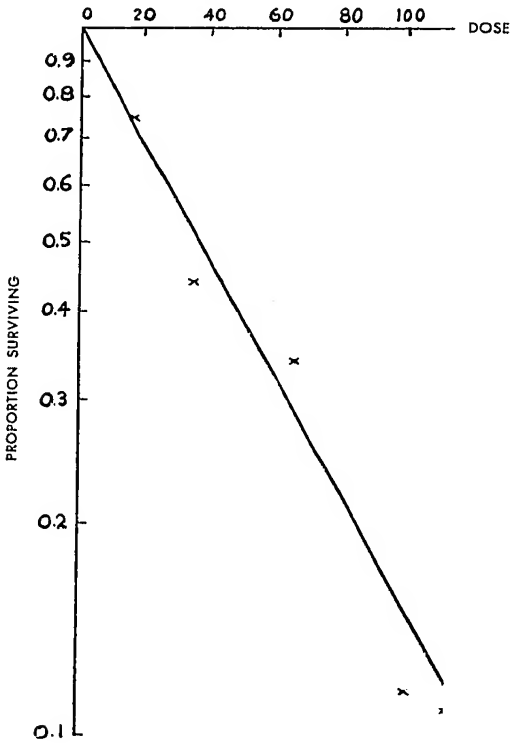


FIG. 1.

Relationship between dose of *B. anthracis* spores and ln proportion surviving of guinea pigs, showing regression line (fitted by eye).

sion lines were obtained in widely different positions. The 8 experiments on *B. typhosum* were picked at random from 56 similar ones.

TABLE II. VARIOUS RELATIONSHIPS FITTING ONE FIXED LINE

Organism	Test Animals	Route of Infection	No. of Expts.	χ^2	Degr. of fr.	Graph
Brucella suis (4)	300 Guinea pigs	Respiratory	4	3.2	6	fig. 2a
B. anthracis (3)	990 Guinea pigs	Respiratory	5	25.9	22	fig. 2b
B. typhosum (1)	640 Mice	Intraperitoneally	8	19.8	14	fig. 2c

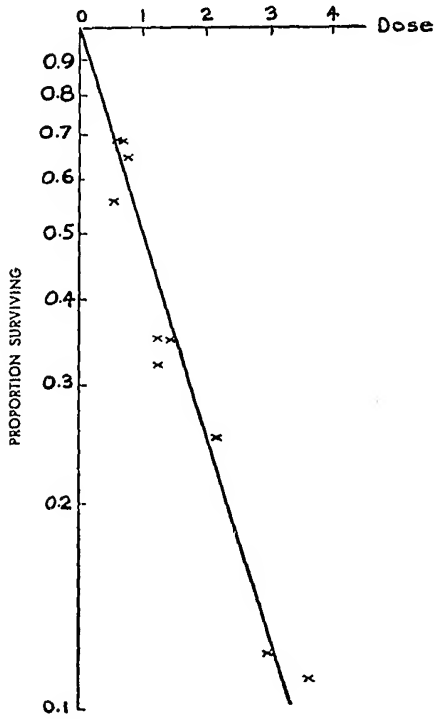


FIG. 2a.

Fixed regression line in relation to the result of 4 experiments with *Brucella suis*; doses of the sub groups are expressed as multiples f of the ED_{50} of each individual experiment.

Estimation of Relative Potency

If different strains of the same pathogenic organism are to be compared as to their virulence with respect to a host, one dose-response line is fitted for each strain and the ratio of the slopes estimates the relative potency. For let p_s and p_u be the slopes of "standard" and "unknown" respectively and n_s and n_u be doses producing a common response S; it follows from equation (1a) that $n_s p_s = n_u p_u$ or

$$\text{Relative Potency } R = \frac{n_s}{n_u} = \frac{p_u}{p_s}$$

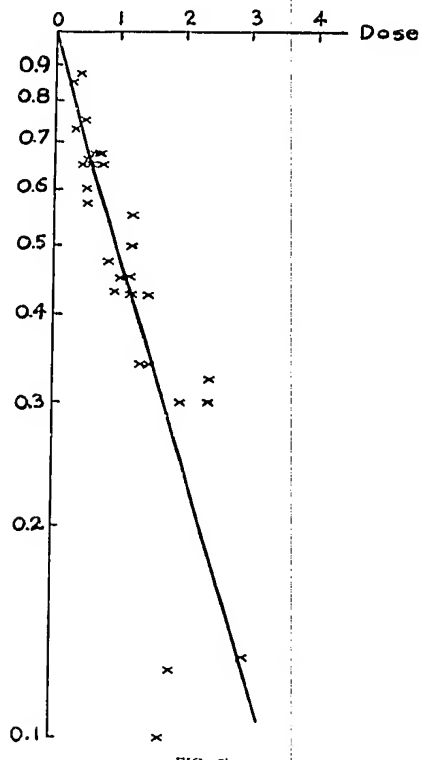


FIG. 2b.

Fixed regression line in relation to the result of 5 experiments with *B. anthracis*.

Since Var (p_u) and Var (p_s) have been computed already together with p_u and p_s, fiducial limits to R are readily assigned as

$$\frac{R \pm \sqrt{g[R^2 + (1 - g)\ \mathrm{Var}\ p_u/\mathrm{Var}\ p_s]}}{1 - g}$$

where $g = t^2\ \mathrm{Var}\ p_s/p_s^2$ and t is the normal deviate for the level of probability to be used (5).

Comparison with probit analysis

This Department used to apply the methods of probit analysis to microbiological dose-response relationships (6). When for example ani-

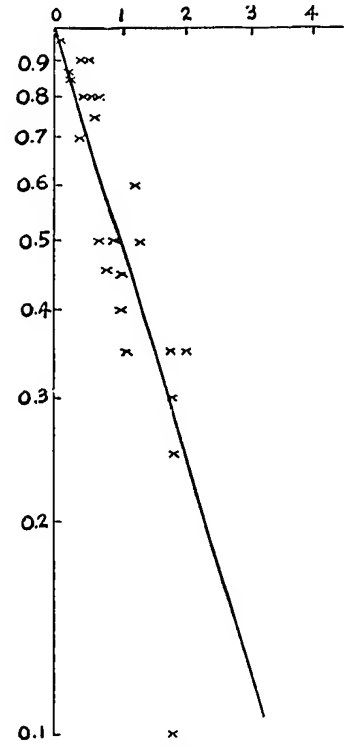


FIG. 2c.

Fixed regression line in relation to the results of 8 experiments with *B. Typhosum*.

mals were exposed to pathogens, a ($\log_{10}$ dose) — (probit killed) regression equation has fitted well the results in most cases. If the hypothesis expressed in this paper holds good, i.e. if $\ln S = -pn$ (equation 1a) actually applies, all the probit lines computed so far ought to show about the same slope within experimental error (Examples are quoted in Ref. 2). For $\log_{10}$ dose — probit killed transforms an ideal dose — ln proportion surviving into a slightly bent curve (see fig. 3) and it is easily proved that, if Y is the Probit, the slope $dY/d \log_{10} n$ at the ED_{50} equals 2 (very nearly)

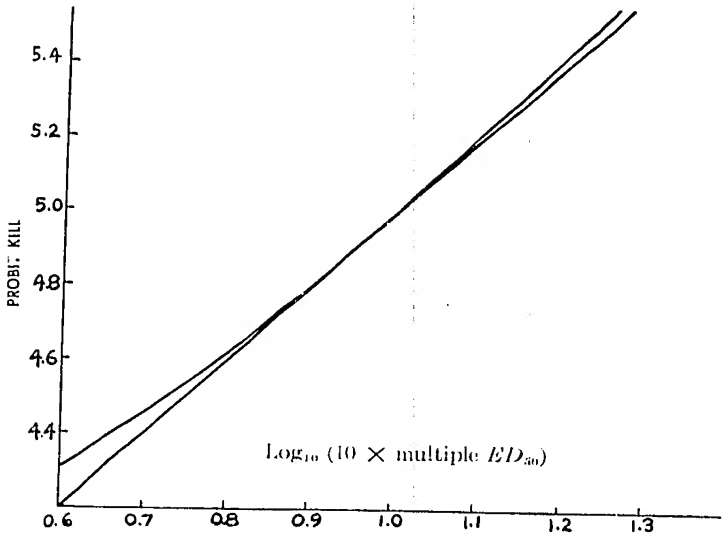


FIG. 3

An ideal dose — ln proportion surviving relationship plotted as probits against log dose. The slope at the ED_{50} is 2.

Consider

$$1 - S = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y-5} e^{-\frac{1}{2}u^2} du \quad \text{and} \quad \ln S = -pn$$

Now

$$\frac{dY}{dS} = -\sqrt{2\pi}\, e^{\frac{1}{2}(Y-5)^2} \qquad (A)$$

$$\frac{dS}{dn} = -Sp \qquad \text{(B) and}$$

$$\frac{dn}{d \log_{10} n} = n \log_e 10 \qquad \text{(C)};$$

Substituting in the product of (A) × (B) × (C) for $Y = 5$, for $S = 0.5$ and for $pn = -\ln 0.5$ gives 2.0003.

It may be concluded that

(i) the two models differ only slightly and could not be experimentally distinguished without forbidding expenditure in test animals,

(ii) that the calculated ED_{50} should be on the whole lower when probit analysis is used (see Table III),

(iii) the present method compares favourably with probit analysis as far as computation labour is concerned.

Examples from practice have been worked out by both methods and the results are given in Tables III and IIIA. The estimates of relative potencies provided by the two techniques are in good agreement throughout; the probit method appears in general to give slightly narrower fiducial belts but the point at issue, of course, is not which method gives more apparent precision but which one is based on the more appropriate mathematical model.

TABLE III. COMPARISON OF TWO METHODS OF ANALYSIS

A. . . . . probit analysis

B. . . . . present method

Organism	χ^2/df.		ED_{50} and its 95% fid. lts.	
	A	B	A	B
B.anthracis, 1μ particles*	0.72	0.41	0.24 → 0.34 → 0.43	0.29 → 0.36 → 0.47
B.anthracis, 3.5μ particles	1.33	0.94	0.29 → 0.36 → 0.44	0.30 → 0.36 → 0.47
B.anthracis, 4.5μ particles	1.16	1.08	0.40 → 0.51 → 0.70	0.44 → 0.53 → 0.68
B.anthracis, 8μ particles	1.46	1.51	3.3 → 3.8 → 4.7	3.3 → 4.0 → 5.1
B.anthracis, 12μ particles	0.10	1.28	4.0 → 5.3 → 7.1	5.1 → 6.0 → 7.3
Brucella suis	0.06	0.77	6.9 → 31.7 → 52.7	34.0 → 45.3 → 68.0
Brucella suis	0.88	0.67	25.3 → 35.3 → 57.0	25.5 → 35.0 → 55.9
Brucella suis	0.26	0.05	21.0 → 34.8 → 47.9	26.0 → 36.1 → 57.0
Brucella suis	0.44	0.44	19.0 → 31.8 → 44.1	24.1 → 32.7 → 51.0

*See Table I.

TABLE IIIA. POTENCIES OF 2 MUTANTS OF Ty_{22}—STRAIN OF *B. Typhosum* RELATIVE TO PARENT STRAIN. 2 DIFFERENT EXPERIMENTS. RESULTS OBTAINED BY THE TWO METHODS

Expt.	Organism	χ^2 df.		Rel. Potency and its 95% fid. lts.	
		A	B	A	B
MVT. 75	Ty_{22}	1.96	3.04		
	Mutant a	0.27	0.23	2.25 →3.39 →4.90	2.17 →3.43 →5.78
	Mutant b	0.19	1.60	0.73 →1.05 →1.56	0.58 →1.01 →1.70
MVT. 85	Ty_{22}	0.95	0.56		
	Mutant a	0.24	1.27	0.02 →0.029→0.042	0.019→0.030→0.049
	Mutant b	0.27	1.47	0.047→0.069→0.099	0.057 →0.075→0.110

Choice of doses

When experimental evidence has shown that the dose — ln proportion surviving regression line holds good for a certain host-parasite relationship, the question arises: What is the most economical investment of test animals? In other words: Which dose n will minimise the variance of the regression coefficient p?

Using equation (4) for one point:

$$\frac{d^2L}{dp^2} = -(m - r)n^2 \frac{e^{pn}}{(e^{pn} - 1)^2}$$

and putting $(m - r)$ equal to its expected value $m(1 - e^{-pn})$ we obtain

$$\frac{-1}{\frac{d^2L}{dp^2}} = \text{Var}\,(p) = \frac{e^{pn} - 1}{mn^2} \tag{7}$$

and hence the Invariance of p can be expressed in the form

$$\frac{1}{\text{Var}\,(p)} = \frac{m}{p^2} \frac{(\ln S)^2\, S}{1 - S} \tag{7a}$$

Differentiating equation (7) w.r. to n and putting

$$\frac{d(\text{Var}\,p)}{dn} = 0,$$

the equation is simplified to

$$e^{pn}(2 - pn) - 2 = 0. \tag{8}$$

Hence Var (p) is a minimum for

$$pn = 1.5936 \tag{9}$$

which corresponds to

$$e^{-1.59} = 20.3\% \text{ survivors} \tag{9a}$$

$$\text{Inv } (p)_{max} = \frac{n}{p^2} \times 0.6476 \tag{9b}$$

In fig. 4 equation (7a) is plotted as a percentage of its maximum value (equation 9b). It will be noted that the efficiency within the range of 10% 35% survivors is hardly affected; in practice the experimentor would aim at a moderately low survivor rate and know that, if he misses within reason the theoretical 20.3% optimum (equation 9a), economy would still not be appreciably impaired.

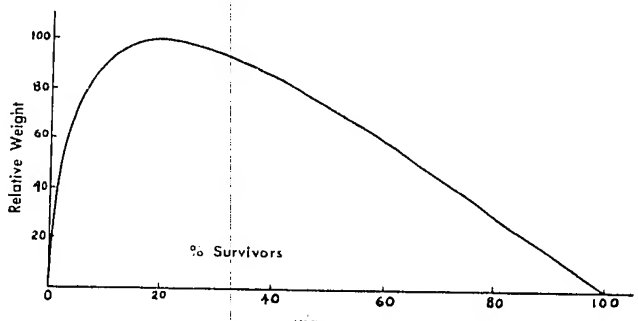


FIG. 4.

Relative Invariance of an estimate of p based on a single dose level as a function of the proportion surviving.

If little is known about p and even linearity is doubtful, obviously more than one point is needed; but high survivor rates which carry little weight will be avoided, if possible.

Application to dilution series

After this paper had been drafted the close analogy with the analysis of dilution series was pointed out to me by Mr. M. J. R. Healy. Using the notation of this paper the relevant problem can be presented in the following rather condensed form. (For details references (7), (8) may be consulted.)

REFERENCES

(1) Bacon, G. A., Burrows, T. W. and Yates, M., The Effects of Biochemical Mutation on the Virulence of *Bacterium Typhosum*. *Brit. J. of Exp. Path. 31*: 714, 1950 and *32*: 85, 1951.
(2) Druett, H. A., Bacterial Invasion, *Nature 170*: 288, 1952.
(3) Druett, H. A., Henderson, D. W., *et al.* in preparation.
(4) Elberg, S. S. and Henderson, D. W., Respiratory Pathogenicity of *Brucella*. *J. of Inf. Dis. 82*: 302, 1948.
(5) Fieller, E. C., A Fundamental Formula in the Statistics of Biological Assay, and some Applications. *Quart. J. Pharm. Pharmacol., 17*: 117, 1944.
(6) Finney, D. J., *Probit Analysis*, Cambridge Univ. Press, 2nd Edn., 1952.
(7) Finney, D. J., *Statistical Method in Biological Assay* (§21.5) London: Charles Griffin & Co.
(8) Fisher, R. A. & Yates, F., *Statistical Tables for Biological, Agricultural and Medical Research*, Edinburgh: Oliver & Boyd.

Presented at the Third International Conference on Biometrics

Bellagio, Italy, September 2, 1953.

FACTORS OF DOSAGE AND HOST DETERMINING ANTIBODY RESPONSE TO SECONDARY ANTIGENIC STIMULUS.*

By

JOHANNES IPSEN, JR., M.D.

Superintendent, Institute of Laboratories, Massachusetts Department of Public Health, Associate Professor of Public Health, Harvard University School of Public Health, 375 South Street, Boston 30, Mass.

*The experiments were conducted under the sponsorship of the Commission on Immunization, Armed Forces Epidemiological Board, and were supported in part under contract with the Office of the Surgeon General, Department of the Army.

A secondary antigen stimulus, according to Glenny and Südmersen(2) is defined as the injection of antigen into an actively immunized animal. The active immunization may be due to natural infection or to artificial immunization with the same specific antigen. The individual may or may not show evidence of such previous treatment by presence of specific antibody in the serum. The effect of a secondary stimulus is a distinctive, higher immune response than that of a primary stimulus. Smaller doses of antigen are necessary to produce a secondary response, than those needed for a primary response.

The primary stimulus, however, is preferred in assays of the relative potency of toxoids, because greater uniformity of response is expected from a non-immune population. In field trials of toxoids in human population samples, one is often faced with a mixture of partly immune and non-immune individuals, so that the interpretation of the response is equivocal. Accurate information of previous immunization is often essential for a reduction of the variability through covariance. The following presentation is an attempt to evaluate the importance of measurable independent variables and their contribution to the variation in secondary response in human beings.

Response is in this work measured by the antibody titer which is one measurement of immune response. Other response measurements such as percentage of survivals

or percentage of skin reactions to antigen or toxin, are less informative since they have the limitations of all-or-non-reactions. A fundamental relationship between antitoxin titer and all-or-non-reactions has been well substantiated, recently by Holt(3), and Cavalli-Sforza(10).

The discussion is limited to the antitoxin titer at a certain time after the antigen injection, which approximates the time of maximum response. The quality of the antitoxin which recently was subjected to intense study by Jerne(6) is also omitted from this discussion, where antitoxin titer simply denotes the highest dilution of a serum which neutralizes a given dose of toxin to a certain end-point reaction. Since the titers are uniformly determined throughout the assay, this study deals with comparison of response, where the absolute value of the response is unimportant. The notation of the titers in Antitoxin Units per ml. (AU) is in conformity with present convention.

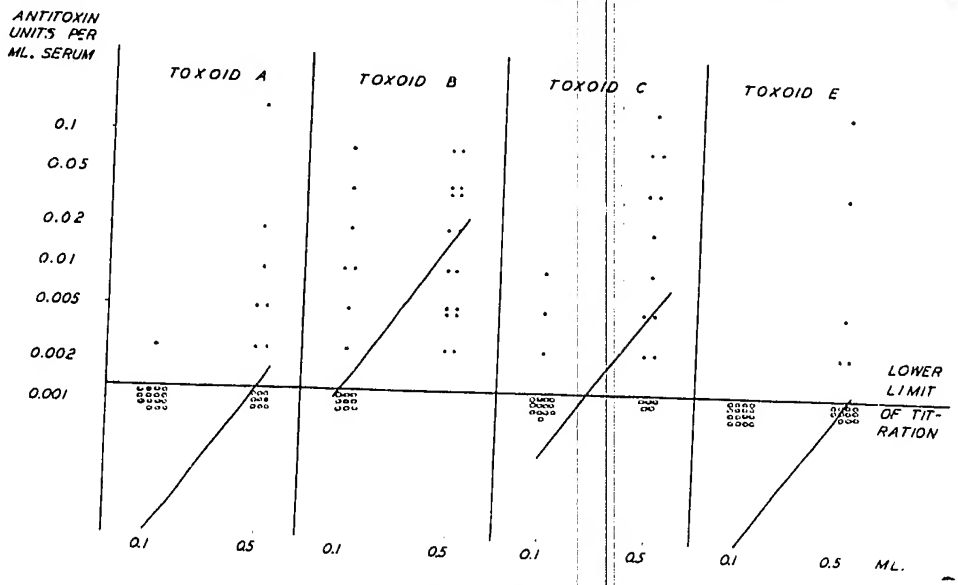
Titrations were performed in mice with constant toxin doses and serial 1:2 distributions of serum.

For a complete analysis of the factors determining a given secondary response the following variables must be included in the experiment and each must be varied over a sufficient wide range.

the first injection, the right hand part are the titers 14 days after the secondary injection).

Relative Potency of Toxoids.

A previous report on these data, (Report[14]), dealt with the estimate of the relative potency of these four toxoids, which had also been assayed in both rabbits and guinea-pigs. The potency estimates were obtained from this assay in humans by construction of reaction curves relating primary response (v) to dose of toxoid.



(Figure 1.)

Reaction curves of four tetanus toxoids in humans (primary response).

The reaction curves clearly places Toxoid B as the most potent, Toxoid C of average potency, and A and E as the least potent. The same potency relationship was also found to hold when the secondary responses were related to the secondary dose.

the dependent variable, and let the above logarithmic relative potencies substitute for x_2 and z, as measurements of the primary and secondary dosage.

Before assumptions can be made as to primary immunizability, a multiple regression analysis was performed with linear and quadratic terms of x_2 and z and their interactions. Only three terms were found significant (see Table 4, Analysis of Variance) so that a prelimary expectancy equation was computed.

$$y' = 7.283 + 0.870x_2 + 1.286z - 0.467x_2z \quad \ldots \qquad (1)$$

on the basis of the matrix of sums of squares and sums of products of deviations. (See Table 3).

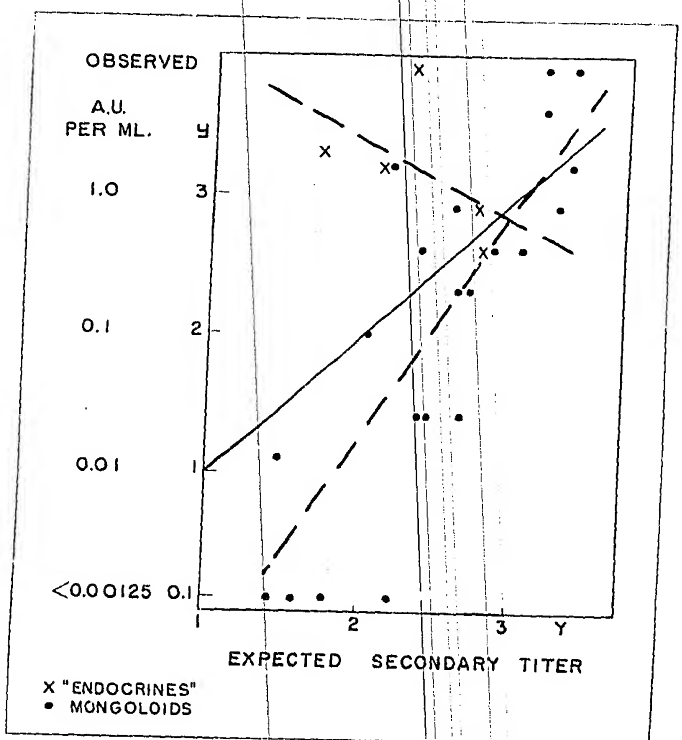
The fit of this expectancy for the overall sample of 128 individuals was next tested for smaller groups. Since the inmates of the institution represented a great variety of constitutions, it was pertinent to investigate whether grouping of the individuals in categories of their somatic and mental diagnosis would reveal any group which deviated in response from the overall expectancy to suggest abnormal immunizability of the group.

Table 5 shows the observed and expected group mean for ten major categories which were classified by the medical staff of the institution, before the immunization program started.

While the observed titers of eight categories, composing 102 persons, agree

well with [illegible] equation (1), there are two groups [illegible] deviate [illegible]-[illegible] of [illegible]. Mongoloids (21 persons) [illegible] low response, and [illegible] [illegible] response.

[illegible] analysis of the deviations from expected values within the [illegible] groups, show that deviations [illegible] largest [illegible] low expectancies, while high expectancies give [illegible] fits.



(Figure 3.)

Observed responses of 21 mongoloids and 5 "dysendocrines" in relation to values expected from equation 1.

[illegible] group [illegible] [illegible] [illegible] [illegible] [illegible] (2). The group "[illegible]" is not a well defined one but since it [illegible] to represent an [illegible] entity [illegible], it is worth including it in an analysis of [illegible].

Since the two abnormal groups have approximately the same [illegible] deviation, the [illegible] can be tested as a linear regression by assigning the factorial x_1 [illegible] to the [illegible], [illegible] to the [illegible], and x_1 [illegible] the remainder [illegible]. The [illegible] of Table 3 [illegible] were also tested, [illegible] the interaction [illegible] without [illegible]. The significance of the factor x_1 and its interaction with [illegible] is brought out in the analysis of variance. ([illegible]).

After the solution of the regression coefficients the [illegible] equation now takes the form:

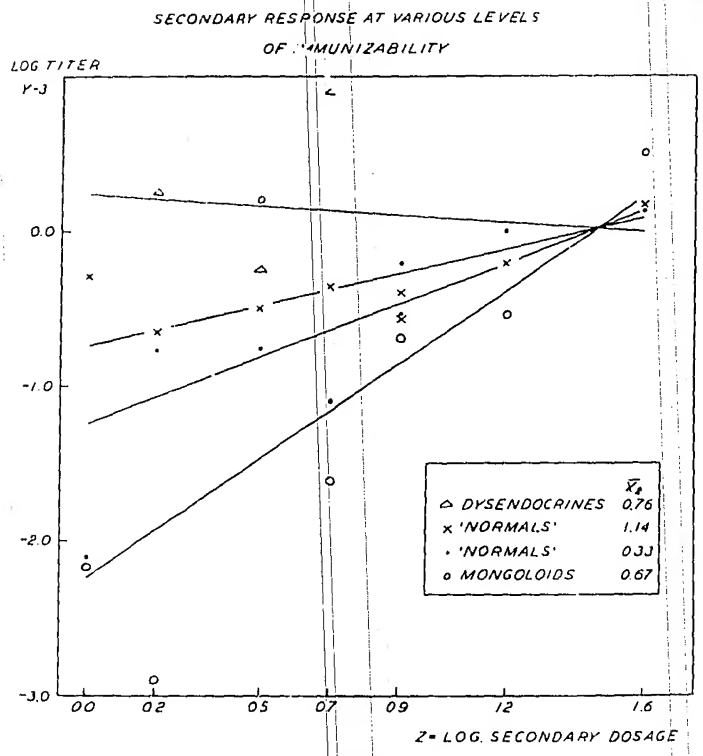
$$y = 1.5[illegible] + 0.63[illegible]x_2 + 1.001z - 0.135x_2z + 1.1[illegible]7x_1 - 0.[illegible]2x_1z \; \ldots \qquad (2)$$

Equation (2) can be rearranged in the form:

$$y = 1.5[illegible] + 0.[illegible]([illegible]x_1 + x_2) + 1.00[illegible]z - 0.135z([illegible]x_1 + x_2) \; .. \qquad (3)$$

which indicates that the inherent factor x_1 has about equal additive effects to the dual functions of the primary dose x_2: one of increasing the secondary response, the other of decreasing the regression of the secondary dose z on the response.

The effect of combination of the primary dose and of the immunizability is clearly shown in Figure 4, which presents the observed data in relation to the expected regression lines of y and z for four groups.



(Figure 4).

Secondary response of the various immunity levels, related to secondary dose.

The upper line represents the expectancy for the 5 dysendocrines who received an average primary dose of $x_2 = 0.76$, which inserted in equation (2) together with $x_1 = 1$, gives

$$y = 3.26 - 0.200z \quad \ldots \qquad (4)$$

[illegible] lowest line represents the equation for the [illegible] (x_2 = −2.[illegible]) to

$$y = 3.[illegible] + 2.162x \ldots \qquad (5)$$

The two intermediate lines are the expectancy for two primary dose levels of the 102 "normals", of which the upper is the expectancy for 47 individuals with relative high primary dosage (x_2 = 1.1[illegible]) and the lowest line represents 5[illegible] persons with a lower primary dose (x_2 = 0.33).

The equations are, respectively

$$y = 2.27[illegible] + 0.[illegible]x \ldots \qquad (6)$$

and

$$y = [illegible] + 0.[illegible]x \ldots \qquad (7)$$

The difference between the two "normal" groups is much less pronounced than the difference between the mongoloids and dysendocrines who had approximately the same primary dosage. The points represent the mean of 1–8 individuals of the same kind and with same secondary dosage.

Difference in Inactivability at Primary Stimulus.

The difference in the two possible groups amount to about 2 x 1.88 = 3.76 in terms of log primary dose as estimated from the secondary response (equation (3)). This means that the mongoloids would need about 5600 times more potent primary stimulus than the dysendocrines

inherent immunizability is not to be included in the primary titer. However, the equation obtained with these 50 individuals has the form:

$$y = 1.61 + 0.72(v + 1.82x_1) + 1.27z - 0.82z(v + 1.31x_1) \quad .. \qquad (7)$$

which is not significantly different from equation (7) obtained from all 128 samples.

The analysis of variance (Table 4) shows that the components due to x_1 are at least as important as those due to v. Hence, it is concluded that inherent immunizability is a function which is independent of the primary response.

Discussion

In the attempt to analyse the variables which influence the secondary response, the procedure has been restricted to analysis of linear components. The true functional equations are probably exponential functions with curves tending toward asymptotes at high values of the respective variables. Such functions are not amenable to explicit regression analysis, and no increase in information was obtained by adding higher terms of a polynomial expansion. The three measurable variables:

(z) potency of secondary dosage

(x_2) potency of primary dosage

(v) antitoxin before secondary immunization

were all found to have some relation to the secondary response. The correlation, however, vanishes at high values of all or either of the variables.

rise in titer after the secondary dose.

3. Individuals with no measurable immunity may be of two categories, previously immunized or non-immunized. These inseparable categories yield widely different results, and immunization histories are necessary for efficient interpretation.

4. Inhomogeneity of the group with regard to immunizability introduces an important variable, which can rarely be determined a priori, except in pure inbred animal strains. This variable is most distinct after secondary stimulus.

References

1. Cavalli, G. Banton, H.J., and Wheeler, R.E.: The antigenicity of single graded doses of diphtheria toxoid in man. Am. J. Hyg. 53, 283-295, 1951.

2. Glenny, A.T., and Sudmersen, H.J. J. Hyg. 20, 176, 1921.

3. Holt, L.B. Quantitative Studies in diphtheria prophylaxis. An attempt to derive a mathematical characterization of the antigenicity of diphtheria prophylactic. Presented at the Third International Conference on Biometry, Bellagio, 2 Sept. 1953.

4. Ipsen, J. Bio-assay of four tetanus toxoids (aluminum precipitated) in mice, guinea-pigs and humans. J. Exp. 78, 426, 1953.

5. Ipsen, J.: Inherent immunizability to tetanus toxoid, based on studies in pure inbred mouse strains. VI International Congress for Microbiology, Rome, Sept. 1953.

6. Jerne, N.K.: Avidity. Acta path. scand. Suppl. 87. Copenhagen 1951.

7. Prigge, R. Die Staatliche Prufung der Diphtherie-impfstoffe und ihre experimentelle Grundlagen. Arb. Staatsinst. exper. Ther. 32, 1935

8. Schoibel, I.F.: Acta path. et microbiol Scandinav. 20, 464, 1943.

9. Siegel, M. Susceptibility of Mongoloids to infection. Antibody response to tetanus toxoid and typhoid vaccine. Am. J. Hyg. 48, 63-72, 1948.

10. Cavalli-Sforza, L.L.: Boll. Ist Sierot, Milanese: 24, 216, 1947.

Acknowledgement

The author wishes to thank Dr. Stanley Raymond and Dr. Wiggin L. Merrill, Wrentham State School for their excellent help in the performance of the field assay, Dr. Huston J. Banton who organized the study; and Mrs. Frieda Kay who performed the antitoxin titration.

Johannes Ipsen: Factors of Dosage and Host Determining Antibody Response to Secondary Antigen Stimulus.

Symbols and Equations.

Y = log (antitoxin titer) plus 3, fourteen days after secondary dose is injected.

v = log (antitoxin titer) plus 3, twenty-eight days after primary dose is injected.

x_1 = factorial to indicate difference in host immunizability (Mongoloids $x_1 = +1$, Dysendocrines $x_1 = +1$, Others $x_1 = 0$).

x_2 = log relative potency of primary dose (see Table 2).

z = log relative potency of secondary dose (see Table 2).

Expectancy equations, for secondary response.

All 128 individuals, disregarding immunizability, including primary dose and secondary dose.

$$Y = 1.243 + 0.870x_2 + 1.286z - 0.667x_2z \quad (1)$$

All 128 individuals, including immunizability.

$$Y = 1.549 + 0.636x_2 + 1.001z - 0.435x_2z + 1.197x_1 - 0.872x_1z \quad (2)$$

Equation (2) rearranged

$$Y = 1.549 + 0.636\,(1.88x_1 + x_2) + 1.001z - 0.435z\,(2.00x_1 + x_2) \quad (3)$$

All 128 individuals, using primary response (v) instead of primary dose (x_2).

$$Y = 1.747 + 0.567\,(2.07x_1 + v) + 0.917z - 0.497z\,(1.74x_1 + v) \quad (8)$$

50 individuals with measurable primary response.

$$Y = 1.61 + 0.72\,(1.80x_1 + v) + 1.27z - 0.82z\,(1.34x_1 + v) \quad (9)$$

TABLE 1

PRIMARY (v) AND SECONDARY TITER (y) (IN LOG ANTITOXIN +3)
FOR 128 INDIVIDUALS, BY PRIMARY AND SECONDARY DOSE OF TETANUS TOXOID.

SEC. DOSE OF TOXOID ml.		z	PRIMARY DOSE OF TOXOID													
			Toxoid A				Toxoid B				Toxoid C				Toxoid E			
			0.1 ml 0.2		0.5 ml 0.9		0.1 ml 0.9		0.5 ml 1.6		0.1 ml 0.5		0.5 ml 1.2		0.1 ml 0.0		0.5 ml 0.7 (x_2)	
			v	y	v	y	v	y	v	y	v	y	v	y	v	y	v	y
A	0.1	0.2	0.1	3.3	0.1	1.7	0.1	2.0	1.0	2.3	0.1	0.1	1.6	3.5	0.1	1.7	0.1	1.0
			0.1	3.2			0.4	2.0	1.6	2.6	0.1	2.0	0.4	3.2	0.1	2.9	0.1	2.6
	0.5	0.9	0.1	2.0	0.1	2.6	0.1	2.3	1.3	3.2	0.1	3.2	2.2	1.7	0.1	1.4	0.1	1.7
			0.1	2.6	0.1	3.5	0.1	1.1	1.0	2.6	0.4	3.2	1.3	3.2	0.1	2.3	0.1	2.9
													0.1	2.9			0.4	2.3
B	0.1	0.9	0.1	3.5	2.2	3.6	0.1	1.1	1.6	2.6	0.1	2.9	1.9	2.3	0.1	2.3	0.1	2.6
			0.1	3.2	0.1	2.9	0.1	1.4	0.7	2.6	0.1	2.9	0.1	2.9	0.1	2.0	1.6	2.9
	0.5	1.6	0.1	3.6	0.1	3.9	1.0	2.9	1.9	3.2	0.1	3.6	0.7	3.2	0.1	2.6	0.4	3.9
			0.1	3.2	0.4	2.9	0.1	2.9	1.3	3.5	0.1	2.9	1.0	3.6	0.1	2.9	0.1	3.2
							0.1	3.2			0.7	3.3						
C	0.1	0.5	0.1	2.9	0.1	2.3	1.6	2.9	1.6	2.6	0.1	3.5	1.6	2.9	0.1	1.1	2.2	2.6
			0.1	1.7	0.7	2.3			0.4	1.7			0.4	3.2	0.1	2.0	0.1	3.2
	0.5	1.2	0.1	3.2	1.0	2.9	0.1	3.2	0.7	2.6	1.0	2.6	1.9	2.0	0.1	2.3	0.1	3.2
			0.4	3.9	0.7	2.9	1.3	2.6	0.7	2.6	0.1	3.2	0.1	2.6	0.1	2.3	0.1	2.6
					0.1	3.5												
E	0.1	0.0	0.1	0.1	1.3	3.2	1.0	3.2	0.7	2.0	0.1	1.4	0.1	2.6	0.1	0.1	0.1	1.1
			0.1	1.1	0.1	2.9	1.9	2.9	1.6	2.3	0.1	0.1	0.1	2.6	0.1	0.7	0.7	2.0
	0.5	0.7	0.1	0.1	0.4	2.9	0.1	1.4	1.9	3.2	0.1	3.9			0.1	0.1	0.1	2.6
			0.1	2.0	0.1	3.5	0.7	2.6	0.4	1.4	0.1	3.5	0.7	1.0	0.1	2.0		

Mongoloids.

Dysendocrines

0.1 indicates values less than lower limit of titration (0.00125 AU).

Table 2

Mean Primary and Secondary Response, by Potency of Toxoid Dose

Toxoid	Dose	Relative Potency	Log Rel. Potency	Number	Primary Response	Number	Secondary Response	Primary Response*
F	0.1 ml	1.0	0.0	16	0.100	16	1.79	0.51
A	0.1 ml	1.6	0.2	16	0.119	15	2.27	0.40
C	0.1 ml	3.2	0.5	16	0.213	14	2.49	0.66
E	0.5 ml	5.0	0.7	16	0.400	14	2.16	0.36
A	0.5 ml	8.0	0.9	16	0.475	18	2.48	0.43
B	0.1 ml	8.0	0.9	16	0.550	16	2.61	0.57
C	0.5 ml	16.0	1.2	16	0.888	17	2.84	0.51
D	0.5 ml	40.0	1.6	16	1.150	18	3.25	0.47

*Mean Titer of Group Before Secondary Dose was Injected.

Table 6

Primary Response in Three Groups of Different Immunizability.

Group	Number	Average Primary Dosage	Percent with Measurable Antitoxin	Mean Primary Response ($\bar{v}$)
"Normals"	102	0.77	41.2%	0.521
Mongoloids	21	0.67	23.8%	0.229
Dysendocrines	5	0.76	60.0%	0.880
Difference in Immunizability Mongoloids vs. Dysendocrines (in terms of primary dosage)			0.35 ± 0.31	0.79 ± 0.40

Table 3.

Matrix of Sums of Squares and

Sums of Products of Deviations from Mean

128 Observations

	z	x2	zx2	x1	zx1	y
z	31.56					24.93
v	0.12	30.46				11.25
zv	42.39	23.04	42.39			22.24
x	-2.95	1.80	-1.82	24.00		12.31
zx1	-8.12	0.38	-5.17	17.68	20.39	-2.24
y						100.2437
Means	0.778	0.750	0.580	-0.125	-0.120	2.507

	z	v	zv	x1	zx1	y
z	31.56					24.91
v	-0.21	46.11				12.79
zv	15.07	34.21	42.42			15.02
x	-2.95	7.39	4.13	24.00		12.31
zx1	-8.12	5.60	2.20	17.68	20.39	-2.24
Means	0.778	0.487	0.377	-0.125	-0.120	2.507

50 Observations (v larger than 0.1)

	z	v	zv	x1	zx1	y
z	12.83					5.03
v	1.59	16.25				0.96
zv	12.07	10.96	24.13			3.88
x	-1.59	3.38	1.01	7.92		1.69
zx1	-3.65	3.81	0.86	7.45	8.39	0.40
y						18.3442
Means	0.806	1.090	0.847	-0.040	-0.064	2.746

Table 1

Analysis of Variance

Primary Dose (x_2), Secondary Dose (z) and Immunizability (x_1)

Component	Sum of Squares	D.F.	Mean Square
Total (y)	100.2437	127	
Regression on z	19.6615	1	
Additional regr. on x_2	4.0898	1	
---- interaction x_2z	3.2442	1	
	73.2482	124	0.5902
Additional regr. on x_1	7.5678	1	
---- interaction x_1z	4.5401	1	
Remainder	61.1403	122	0.5011

Primary Response (v), z and x_1

Total (y)	100.2437	127	
Regression on z	19.6615	1	
Additional reg. on v	3.6398	1	
---- interaction vz	4.2811	1	
	72.6613	124	0.5860
Additional regr. on x_1	6.5125	1	
---- interaction x_1z	4.9122	1	
Remainder	61.2366	122	0.5019

50 Individuals Having v Larger Than 0.1

Total (y)	18.3442	49	
Regression on z	1.9693	1)	
Additional regr. on v	.1566	1)	
---- interaction vz	1.3829	1)	
Additional regr. on x_1	.5786	1)	1.1320
---- interaction x_1z	1.5727	1)	
Remainder	12.6841	44	0.2883

Table 5

Observed and Expected Means for Ten Diagnostic Categories. Expectation from Equation (1), Disregarding Inherent Immunizability, and from Equation (2), Including Group Immunizability

Diagnostic Category	Number n	Observed Mean	Equation (1) Expected Mean	Equation (1) "Student's" t^*	Equation (2) Expected Mean	Equation (2) "Student's" t^{**}
[illegible] SPECIFICITY	44	2.49	2.50	−0.04	2.54	−0.43
FAMIL. AL. LEUKONS	32	2.65	2.55	+0.70	2.59	+0.43
CENTRAL NERV. SYST. MALFORMATIONS	5	2.90	2.76	+0.41	2.77	+0.41
EPILEPSY AND OTHER DEFECTS	6	2.60	2.27	+1.03	2.36	+0.83
PSYCHOSIS AND MENTAL DEFICIENCY	6	2.20	2.52	−1.03	2.56	−1.26
POST-TRAUM. IDIOCY	6	2.82	2.66	+0.49	2.72	+0.32
POST-INFECT. IDIOCY	4	2.38	2.49	−0.30	2.55	−0.49
	102	2.56	2.52	+0.51	2.56	−0.09
MONGOLOIDS	21	2.08	2.49	−2.04	2.07	+0.05
DYSENDOCRINES	5	3.18	2.23	+2.52	3.15	+0.13

$^*\chi^2 = \frac{(\text{Obs}-\text{Exp})^2 n}{0.5907}$ $^{**}\chi^2 = \frac{(\text{Obs}-\text{Exp})^2 n}{0.5071}$